平成 22 年度　国立情報学研究所　市民講座　第１回
「多言語世界の扉を開く翻訳技術　−人間の翻訳と機械の翻訳は何が違うのか？−」
講師：　影浦　峡　（東京大学教授）

◆　市民講座の紹介

本日は国立情報学研究所の市民講座にお越しいただき、ありがとうございます。
本日、コーディネーター、司会という役目を担当しております、国立情報学研究所情報学プリンシプル研究系准教授の稲邑と申します。
では、国立情報学研究所の市民講座を簡単に紹介させていただきます。
この市民講座は平成 15 年より情報学に関連したテーマを国立情報学研究所の教員が、皆さま一般の方向けへの公開講座として開設されました。年 8 回講演が行われます。本日はその 1 回目です。
そして、終了した講義の映像、資料、質問への回答は、国立情報学研究所のホームページに公開されることになっております。
本日は私以外にも国立情報学研究所コンテンツ科学研究系の准教授相原 健郎もコーディネーターとして参加しています。
今日は今までとは違って初めて聴覚障がい者のために手話通訳と、パソコン通訳を行うことになりました。

◆　講師の紹介

それでは、本日の市民講座の講師の方をご紹介します。
今日の講師は、影浦 峡（かげうら・きょう）先生、東京大学大学院教育学研究科教授の方です。
影浦先生は 2001 年 4 月から国立情報学研究所の助教授として着任され、2005 年 10 月に東京大学大学院教育学研究科准教授に異動され、2009 年から東大の大学院の教授でいらっしゃいます。
影浦先生は、プロフィールにあるように、日本図書館情報学会日本図書館情報学会賞、あるいは言語処理学会の第 12 回年次大会優秀賞などの受賞歴に加え、NPO 法人森林サポーターズ・クラブ埼玉県民の森きのこ観察会きのこクイズ・トップ賞という、非常にユニークなご経歴をお持ちの先生です。
今日の市民講座のタイトルは、「多言語世界の扉を開く翻訳技術　人間の翻訳と機械の翻訳は何が違うのか？」ということで、講演をお願いしたいと思います。
では、影浦先生、よろしくお願いします。

・・・

ます。その関係で、多言語世界の翻訳技術という話を今日はします。

・スライド２
「お話の概要」
まず、話の概要です。だいたい６点話します。
第１点目は、翻訳とはそもそもどんな行為か。ここでは問題提起だけします。答えは後ほど。
それを一応踏まえた上で一旦脇に置いて、次に機械の翻訳メカニズムと現状について。
一部の方はこれが一番ご関心があると思います。
ただし、非常に単純に、本当にアイデアだけご紹介します。
技術的な詳細には踏み込みません。
それをさらに踏まえて、そもそも翻訳とはどんな行為かということに戻ってきます。機械でやっている翻訳と、人間でやる翻訳とはどう違うか、という差異を理解しながら、では、人間のやっている翻訳はどんなものかをはっきりさせたいと思います。
その後は、少し現在未来志向の話をして、機械翻訳が仮にもっと向上したら、それでどのくらい使えるようになるか、という話。
仮にもっと向上して、使えるようになると言っても、どうしても使えないところがあるとすると、それは技術ではなく、何の問題なのか。そこでコミュニケーションの原理にかかわるお話に入ります。
最後に、では、実際に人間と機械の翻訳というものを、社会の中でどう使っていくかという話をしたいと思います。

・スライド３
「翻訳とはそもそもどんな行為か（プラトニックに）（１）」
まず、翻訳とはそもそもどんな行為か。早速、次の文章を訳してみましょう。ハングルで書かれていますね。最近ちょっと勉強していますが。「ナヌン　デハゲ　クンムハムニダ」。
だいたいあっていますか？　間違っています？　これ、分かる方？
次にいきます。
Jo treballo per a una universitat.
これが何語か分かる方？

会場／スペイン語？

影浦／スペイン語だとトラバホ…だと思います。
これは、カタロニア語なんですね。
1975 年まで、スペインのフランコ独裁政権下で抑圧されていた言語で、これに関しては『サルバドールの朝』という有名な映画が日本でも公開され、その小説版も出ています。

次、もう１つ。
Mimi kazi kwa chuo kikuu cha.

これ、何語か分かる方？
これは、スワヒリ語。

これはどうでしょう。
I work at a university.
これ、分かる方？いらっしゃいますか？

会場／英語？米語？

影浦／残念でした。これはオーストラリア英語なんです。発音が微妙に違います。
失礼しました。見かけではわからないですね。

・スライド４
「翻訳とはそもそもどんな行為か（プラトニックに）（１）」
次にこれを訳せた人？
たぶん、わかる方は大体訳せたと思います。
1 番目訳せた方？　2 番目は？
これはスペイン語などと一緒なので、見れば大体わかりますね。
特にカタロニア語を知らなくてもこれを訳せる方は、いらっしゃる。スワヒリ語を訳せた方は？
300 人だとなかなかスワヒリ語をカバーするには至らない？
最後はオーストラリア英語なのでちょっと微妙です。「大学」と訳していいのか、
オーストラリアに大学はあるのか？　ちょっと微妙ですが、まあ、大体訳せる。

・スライド５
「翻訳とはそもそもどんな行為か（プラトニックに）（１）」
次にこれを訳してみましょう。
“States Parties shall assure to everyone within their jurisdiction effective protection and remedies, through the competent national tribunals and other State institutions, against any acts of racial discrimination which violate his human rights and fundamental freedoms contrary to this Convention, as well as the right to seek from such tribunals just and adequate reparation of satisfaction for any damage suffered as a result of such discrimination.”

これ、訳せる方？
そこそこいらっしゃる可能性があるかと思います。

・スライド６
「翻訳とはそもそもどんな行為か（プラトニックに）（１）」
まあ、先ほどの英語やカタロニア語について皆さんに「訳せましたか？」と聞いて訳せる

方、英語はかなりの数いらっしゃいましたが、皆さんが今、訳した行為は翻訳なのか？
というのを考えることが、最初の「翻訳とは何か、プラトニックに」の問題提起です。
その時に日本語には便利な言葉があり、中学校・高校では、英文和訳とか仏文和訳などの言葉がありますが、それと翻訳の違いは何か。
いきなり反発をくうことを覚悟で、大雑把に言うと、言語学者さんが特に反発しますが、ヨーロッパ、カナダでは、英語学校と翻訳学校は反目してます。文学があると三者三様でまた仲が悪い。
全然関係ないのですが・・・。

英文和訳とかＸ文Ｙ訳というのは、常識的意味で言語に関わることである。
それに対して翻訳をどう特徴づけるか？
仮に英文和訳と差別化して英和翻訳を特徴づけるとすると、あえて言うといささかも言語に関わる行為ではない、これが翻訳です。
正確に言うためには、言葉、言語をどう扱うか明らかにしなくてはなりません。
よく言われますが、バイリンガルだから翻訳ができるとは限らない、というのと関係します。
常識的に言語学の扱う範囲でという話をしました。
言語とは何か、規定してみます。常識的には、言語とは日本語、英語、スワヒリ語、カタロニア語などがあり、そういうもので、世界では3000から1万いくつの言語があります。
ところが言語学が扱う意味での言語、潜在的に無限の的確文を生成する有限の語彙、及び文法などというものあるいは、「私たちは日本語を話す」という場合に、その日本語なるものを見た人は誰もいない。
我々が見たことがあるのは、表現されたものだけです。
表現されたものは、歴史上どんなに数えても有限ですから、数え上げれば言語学の役割はおしまいか、そうすると言語学は言語についての博物学なのか、いやそうではない　というのがチョムスキーさんが言ったことで、そうではないとするなら、逆にいうと言語表現、テキスト、メディアに記録され、流通されるものとして私達が目にしている言語と、それから言語学でいわれ、そしてさらに私たちがここで暫定的に多くの人が日本語を話しているといったときの日本語というものとはずいぶん違いがあるということです。
ここで第 1 のトピックを暫定的におしまいにして、次に機械の翻訳メカニズムと現状に入ります。

・スライド７
「お話の概要」

・スライド８
「機械の「翻訳」メカニズムと現状」
先ほどの４つの言語についての文章を、機械翻訳にかけてみました。
（G）は何の略でしょう？Google 翻訳の（G）です。
E は Excite 翻訳の（E）です。

個人的にはExcite翻訳は気に入ってます。
まず、ナヌン　デハゲ　クンムハムニダを Google 翻訳にかけると、「私は大学に勤務しています」という訳になります。
Excite翻訳にかけると、「私は大学に勤めます。」です。
微妙に意味が違う。
勤務しているというと、今勤めている。勤めますだと、これから。
そこを除けば、大体どちらもちゃんとしてます。

次にJo treballo per a una universitat.ですが、Google翻訳は「私は大学のために働く」です。
一方Excite翻訳はカタロニア語は扱っていないので、カタロニア語の翻訳はありません。

３番目。
Mimi kazi kwa chuo kikuu cha.
スワヒリ語は動詞、名詞、形容詞の数を合わせますが、これも「私は大学のために働く」とGoogle翻訳では出てくる。
Excite翻訳はスワヒリ語、日本語の翻訳対は扱っていないので、入手できませんでした。
英語は不思議なことにGoogleとExciteが同じ訳を出していて、「私は大学で働いています」です。

これくらいの訳だと、人間が見て分かる程度、そして十分自然な程度に機械翻訳は翻訳しています。
次に、これを訳してみましょう。

・スライド５
“States Parties shall assure to everyone within their jurisdiction effective protection and remedies, through the competent national tribunals and other State institutions, against any acts of racial discrimination which violate his human rights and fundamental freedoms contrary to this Convention, as well as the right to seek from such tribunals just and adequate reparation of satisfaction for any damage suffered as a result of such discrimination.”

今、私のところにリーズ大学の翻訳研究所の先生が１年滞在しています。
英語ネイティブでもさっと読んで意味が分かるのは大学くらいの教育を受けた人。
でないと、さっと読んだだけでは意味が分からないらしいです。
機械翻訳にかけました。
まずはGoogleさん。

・スライド９
「機械の「翻訳」メカニズムと現状」

「締約国は」国際法では、この訳は正しいですね。
締約国は、すべての人に
・・・となります。
・・・ちょっといかがなものか、と思うわけですが。
次に、Excite 翻訳。

・スライド 10
「機械の「翻訳」メカニズムと現状」
Excite 翻訳は私はお気に入りなので、締約国の訳に失敗していたからといって怒らないでください。
州の党は・・・となります。
Google 翻訳とはだいぶ違うかなと思ったんですが、
最後の文学的ニュアンスが似ています。
「・・・保証しなければならないただ、十分な補償や満足度など法定損害そのような差別の結果として被ったから。」
これは今日の話とは関係ないですが・・・。

・スライド 11
「機械「翻訳」メカニズムと現状」
以上から機械翻訳について分かったことを簡単に整理してみます。

まず、スワヒリ語も含めておそらく、人間が 300 人ぐらい、こういう場所に集まったというよりは、扱う言語の範囲は広い。
Google は数十から 100 言語くらい。
もう 1 つは最初の 4 例について考えます。
そこそこ簡単な文は訳せる。それが 2 点目にわかることです。
3 点目、難しくなると支離滅裂になる。
私たちは難しくなったら支離滅裂にならないでしょうか。
論理的にすぐさま訳せと言われて、スラスラできる人は少ない。でも Google も Excite も、数秒でやってくれる。
皆さん支離滅裂にならずに訳してください、と言ったときに、訳せないと思います。
他にどんなことがわかるか、それを考えていくのが、次の課題になります。
ここまでで、翻訳とは何かという問題提起だけしました。

次に、機械翻訳の現状を例に基づいて見てみました。

ここから何がわかるか、更に展開する前に技術的なところだけ、簡単に、本当に一般的な事を、多少の単純化するための嘘偽りをあえて避けずに、紹介します。
機械翻訳の手法というと、規則に基づくものと、用例に基づくものと、統計的な機械翻訳があります。

おおざっぱに言うと Excite 翻訳は規則に基づく翻訳、Google は統計的な機械翻訳を使っています。

・スライド 12

「機械の「翻訳」メカニズムと現状：規則に基づく翻訳」

この 3 つについて、簡単に技術的なところを説明します。

まず、規則に基づく翻訳。2 つの言語の辞書と文法規則と対訳辞書を準備します。

日本語と英語なら、英語の辞書と文法規則、日本語の辞書と文法規則、さらに日本語と英語の対訳辞書。

それに対してどうするか。

日英翻訳なら、「わたしはおじさんです」という文があったときに、辞書と文法規則で構文を解析します。これは構文解析してなく、単に単位分割しているだけなのですが、実際には、構造を解析します。私、は、おじさん、ですのように。

さらにこれを言語に依存しない表現に変えます。

一応、暫定的にどの言語でも主部と述部に文はわかれます。

あるいは、文は主語と述語と目的語、間接目的語などにわかれると言われています。

そんな情報にわけています。主語は「私」で、おじさんは述部。

ここで、対訳辞書を適用します。

順番はインチキですが。

そうすると「私」は「I」になり、「おじさん」は「ojisan」になります。

その後は、英語側の文法規則を使い、ここから、この言語に依存しない表現から英語の文章を規則的につくる。

そうすると、I am an ojisan という英語の文が出てきます。

・スライド 13

「機械の「翻訳」メカニズムと現状：規則に基づく翻訳」

こんなときに困ることは何か。

例えば形容詞は名詞にかかるという日本語の規則で、「太った山本さんの猫」というときに太っているのは、山本さんか、猫かという問題が出る。

解析の時に、太った山本さんのネコは、太った猫なのか、太った山本さんなのか、曖昧性が発生します。構文規則だけでは。

さらに前置詞句は述語に係ったり、名詞にかかったりする 2 つの規則があるので、

He saw a woman with a telescope. と言うと、望遠鏡でのぞき見をしたのか、

女性が望遠鏡を持っていたのか、これが俗に構文的曖昧性といわれています。

私たちの日常会話では、このような例はふんだんにあります。

通常私たちはコミュニケーションの中で普通に意味からその辺を類推していると言われています。

さらにわからなかったら聞き返す。

とりわけ親密になればなるほど「あれ取って」「それ？」とか。

もう１つ構文的曖昧性の他に意味の曖昧性があります。
He went to the bank and got the money.と He went to the bank and picked up morels.と言ったときに、morels というのはアミガサタケです。
生のままだと微妙に毒ですが、乾燥させて水で戻して食べます。
とりわけ大陸ヨーロッパで人気のある春に生えるキノコです。
乾燥品はトリュフに次いで高い値段がつきますから、フランスやスペインのマーケットでは乾燥品が売られています。
前者は、銀行に行ってお金をおろす、後者は川べりに行ってアミガサタケをとった。
川べりといっても東京の川べりはよくわかりませんが、四ツ谷の上智大学の横の中央線沿いのサクラの樹下に４月にこのアミガサタケが生えます。来年、行ってみてください。
・・・全然関係ないことですけど・・。
これは構文的な情報だけだと、絶対に銀行と川岸は区別できないということになります。
でもさっきの Excite 翻訳や Google 翻訳でやってみると、けっこう区別が出来ます。
なぜかというと Excite 翻訳の場合には bank と money が一緒に現れたら、bank はなるべく銀行という意味で表そうという規則がある。一方 Google 翻訳については、後でお話しします。

・スライド 14
「機械の「翻訳」メカニズムと現状：用例に基づく翻訳」
次に統計翻訳の前に用例に基づく翻訳を簡単に解説します。
大雑把にいうと、２つの言語のたくさんの対訳用例と辞書を用意します。
「私はおじさんです」という入力文があったときに、その対訳用例の中に
「私は銀行員です」があった。
おじさんと銀行員が違うので、銀行員のところだけをかえて、banker と銀行員だけ変える。
banker に相当するところを ojisan に書き換えて、I am an ojisan としている。
説明がちょっとインチキですがこのように用例に基づいてパターンの中で違っているところを見つけ出して、過去のパターンを活用しながら訳そうというのが、用例に基づく翻訳。

・スライド 15
「機械の「翻訳」メカニズムと現状：用例に基づく翻訳」
困ることは、言語データは稀なので、稀というのは、今、私が話しているような言語、文というのは、おそらく日本に 1 億何千万という人口がいて、毎日いろいろしゃべっているにも関わらず、まさに今、私が話している文は、誰も話したことがないかもしれない。
その意味では極めて内容がないけど、オリジナルな文です。
つまりそのような過去の用例をどんなにためても、長い単位では当たりにくい。
単語だけで数十万から数百万あるので、単語の並びだと、日本語の場合は前から形容詞がかかるので、名詞のうしろにすぐ形容詞がきて、その形容詞が名詞をかけるということは無いので、単純に 10 万語があったときに、2 つの単語が並んだら、10 万かける 10 万のパターンがあるのではなくて、それよりはるかに組合せのパターンは減ることは減りますが、それでもものすごく多くなります。

ですから、用例ベースの翻訳はなかなか当たりにくい。
これがそこそこ有効に使えるのは、天気予報の翻訳とか。
割と領域と言葉遣いと使われる言葉が限られているところでは結構使えるかもしれません。

・スライド 16
「機械の「翻訳」メカニズムと現状：統計翻訳」
次に統計翻訳。
Google 翻訳がそうですが、2 つの言語のたくさんの対訳用例と、それぞれの言語の沢山のデータと辞書から各言語でのつながり及び対応の確率を計算したものです。
英語の場合に of の次には、大体、the がきます。
特権的に the です。私などは・・
そういう確率をたくさん計算している。
これが日本側と英語側でやってみます。
これが翻訳の問題ではなく、一言語について、どんな単語がどんな品詞と結びつくかですから、それを、確率的な言語モデルといいます。
それに対して、対応の確率、英語でこうだったとき、日本語でどうだったかを探しておいて、それを計算していきます、これを翻訳モデルといいます。
例えば、先ほどの States Parties は締約国と訳される確率が高いと Google が情報を持っている。
困ることは何か。
1 つの言語のつながりの確率を計算するとき、規則に基づく翻訳では、係り受け関係を誠実に文法規則によって計算するんですが、統計的モデルでは、基本的には規則に基づくモデルほど、深く係り受け
関係を解析しません。
A という単語の次に B という単語、そして C という単語がくるというのを 3 つとか 4 つなどの連接で考えるのが基本。
最近はその 2 つが融合してきていて、技術的には発展していますが、今のところはそんな感じです。
先ほどの「太った花子の猫」は、太った猫なのか、太った花子なのか。太った猫ととるには真ん中を越えて係受け関係を見なくてはならない。
そういうのはこれまでの統計翻訳では苦手でした。
その結果、何が起こるかというと、構造が異なる言語では特にそうですが、ちょっと複雑になると構文がむちゃくちゃになることがあります。
一般的にはある程度単純な文より、ほんのちょっと複雑な文について統計翻訳は訳語の選択には極めて有効のようです。
とりわけ Google のようにきわめて大規模な言語表現をデータベースとして持っていて、そこから翻訳エンジンを作るとき、訳語選択は良好であることは多いけれども、ちょっとひねった文だとこけてしまう。
もちろん、Excite 翻訳で翻訳結果が Google よりも優位に文の構造がうまく写し取れていることは、観察されないので、あのぐらい複雑だとどっちにしてもこけてしまう。

以上が非常に簡単な機械翻訳の技術的枠組みの紹介です。

・スライド 17
「機械の「翻訳」メカニズムと現状」
以上から分かることに戻ります。
先ほどまで確認したのは、300 人よりは使う言語の範囲は広い。
簡単な文は訳せそう。難しくなると支離滅裂になる。ただし早い。
次に、ここから分かることは、いずれにしても機械翻訳は言語に関わるプロセスであるということ。
第 1 ブロックで説明した言語学で言うところの言語と言語表現というものの違いに戻ります。
もちろん用例翻訳や統計翻訳は実際に語られたデータを使いますから、その意味では言語に関わるものではないのではないかという可能性もありますが、一旦、規則として抽象化してしまいます。

用例翻訳は別ですが、この場合も一般的用例として抽象化します。
そういう意味では、機械翻訳は言語学の言うところの言語にかかわるプロセスと言えます。

・スライド 18
「機械の「翻訳」メカニズムと現状」
ここで、質問です。
中学生の方いますか？
小学生の方？
研究者？
翻訳のプロ？
日本語を母語としていない方？
いろいろいらっしゃると思います。
Google 翻訳と Excite 翻訳で同じ文を翻訳して比べてみましょう。
小学生が Google で翻訳しても中学生がしても研究者がやっても翻訳のプロが翻訳しても……
日本語を母語としていない人が翻訳しても……。
私のご近所のアライさんのイヌが Google 翻訳しても同じ結果が出ます。
比べてみてください。
ということは、どういうことか。
相手に応じて訳を変えることが出来ない、威厳を持って誰に対しても同じパターンでしゃべるのです、という特徴が機械翻訳にはあることが分かる。

・スライド 19
「機械の「翻訳」メカニズムと現状」
ここまでで、機械翻訳までが終わりです。

機械翻訳の特徴として、最初、翻訳とはそもそもどういうことかで対比した言語的なプロセスといささかも言語にかかわるプロセスではないと言い切ってしまったところとの対比があらためて出てきます。

・スライド 20
「お話の概要」
次に、翻訳とはそもそもどんな行為か（プラトニックに）(2) ということで、暫定的に、じゃあ、人間がやっている翻訳って何なのでしょう？
先ほどプロの翻訳者として手を挙げた人がいるので、怖いのですが、文句言わないで下さい。

・スライド 21
「翻訳とはそもそもどんな行為か（プラトニックに）（２）」
次の例を考えます。
“States Parties shall assure to everyone within their jurisdiction effective protection and remedies, through the competent national tribunals and other State institutions, against any acts of racial discrimination which violate his human rights and fundamental freedoms contrary to this Convention, as well as the right to seek from such tribunals just and adequate reparation of satisfaction for any damage suffered as a result of such discrimination.”

この文。
3 回目ですね、これを読むのはだいぶ、うまくなりました。

・スライド 22
「翻訳とはそもそもどんな行為か（プラトニックに）（２）」
質問です。
本当のプロはこの文を訳す前にどうするでしょう？
ひっかけ問題です。
こんな場所で、プロの翻訳者、手を挙げてくださいと言ったら、この限りでは翻訳できません、というのが返ってくる答えだと思います。
ここで言語自慢の人は無理やり訳そうとしますが、そういう人はだいたいプロではないと思います。
言語自慢が無理矢理訳そうとすることと、プロがこの限りで翻訳できないと言う、
その違いは何か？
その前に、プロの翻訳者はフリーランスだったり、企業内で働いていたりして、会社が受けてきたものをやれといわれ、やらざるを得ない。
この文章を含むものの翻訳を受けたらどうするか？
どうする？　簡単です、訳なんてしません、いわゆる英文和訳しません。
これは人種差別撤廃条約ですから、日本は第 4 条に保留をつけた上で批准しています。

外務省かなにかに正訳がころがっているはず。
だから、正訳を探すのが正しいまっとうなプロの翻訳者のとるべき道です。
そこから遡ると、これが条約であり、何条約かを探す。
専門分野で国際法の翻訳についてのプロという場合は、頭に入ってないといけないし、
一般的プロなら、条約ぽいと思って探せなければプロではない。
次に“対象言語がこの文書と関連づける言語表現域を確定”・・読んでいて分からないのですが、つまりこれを翻訳して読み手はどういう層なのかを考えます。
読み手が普通にお堅い官僚とか企業に勤める人なら、正訳をそのままコピペして使うのが正しい。
新聞や大学ではコピペをしてはいけませんと言われますが、プロの翻訳者はどれだけコピペできるかが、大きなポイントの一つです。
ただし、最近、産業構造的には、翻訳メモリーという翻訳者を支援するシステムが出来ていて、それに対する課金体系が、翻訳メモリーの中であたったものは、翻訳料をとってはいけないという契約が多いのです。つまり、コピペ文は翻訳料をもらえない。
でも、そんなとんでもない話はなくて、プロならプロであるほど、きちんとコピペを使えないといけないから、本当はその専門知識についてもお金を取るべく、労働組合を作って運動しなくてはなりません。
さて、正訳をそのままつかうか対象読者に応じてあえて表現を変えるか。
例えば世界人権宣言の場合。
法務省かな、外務省でしょうか、そこにいくと、官僚の法律的な正訳があります。
その他にはアムネスティインターナショナルのホームページにいくとある谷川俊太郎訳は、一般の人に伝わるようにことばづかいを考えて訳されています。
正訳があるからと言ってそのまま使うのが正しい翻訳であるとは限らない。
それを考えることができないといけない。
言語に関わる行為ではなく、言語表現にかかわる行為というのは、もう一度、元にもどると、翻訳するものというのは「日本語」ではないのです。
日本語から英語への翻訳というのは、「日本語」を翻訳するのではなく、「日本語で表現されたもの」を翻訳するのです。英語から日本語の翻訳というのは、「英語」は誰も見たことがないので英語で表現されたものを翻訳する。
誰かがどこかで発話しないかぎり、翻訳対象となるものは存在しない。
誰かがどこかで発話するというのはどういうことか。誰かがどこかで書くということはどういうことか。
とりわけ、翻訳は書かれた言葉に対してなされるので、誰かがどこかで書いて出版するというのはどういうことか・・というと、歴史的、社会的な環境の中で、ある特有の表現を生み出すということです。そこから歴史的社会的背景から剥奪して抽象化したら言語になりますが、翻訳は言語表現にかかわるものなので、その個々の表現を変換する、訳してやるというのが翻訳です。
ここで結局翻訳は歴史社会的な環境を剥奪するのではなく、歴史社会的な相のもとで扱う。
さらに翻訳結果というのは、私は、たくさんの翻訳をしていて、非常に多くのものは自分だけにやにや喜んでいるだけで、本当に利用者に届けていないのですが、通常、翻訳結果

は利用者に届ける。
そういう意味では利用の側の歴史的、社会的条件も考慮し、それに束縛されることになります。
無茶苦茶、文字が多いですが、私がスライドを書くとこうなります。

・スライド 24
「翻訳とはそもそもどんな行為か（プラトニックに）(2)」
翻訳は、あえて関連する他の行為と差異化する点（つまりあえて関連する他の行為として、英文和訳などから差異化する観点から見るならば)、いかなる意味でも、いささかも認知的あるいは言語的行為ではなく、テクスト/言語表現を、社会に流通可能かつ蓄積参照可能なかたちで提出する、徹頭徹尾社会的な行為である。というのが、まあとりあえず翻訳を特徴付けるためにいささか極端ではありますが、言ってみました。
　言語学的な意味での言語も意味も「翻訳」がそれ以外の何でもなくまさに「翻訳」であることにはいささかも関わりはしない。　そこまでだったら英文和訳とよびましょう。
翻訳元のテクストは、その言語において蓄積されたテキストの織物の中に・・・
（これはロランバルトの表現ですね。こういう感性的な表現を人文学の人はよくしますが、)
位置づけられるべく公表され社会的に蓄積記憶される、翻訳は、翻訳元のテクストが原言語におけるテクストの織物に占める位置に相当する対象言語におけるテクストの織物上の位置に翻訳テクストを位置づける行為である。
読んでいて息がつまりますが、どういうことかというと、大体こんなことになります。

・スライド 25
「翻訳とはそもそもどんな行為か（プラトニックに）（２)」
まず、誤った翻訳行為の図式は、原言語のテクストがあって、中心には言語があり、語彙とか文法規則があるが、基本的には抽象的な言語であり、それを変換すると対象言語のテクストが出てくる。
これがさきほどの機械翻訳の典型的なモデルです。
用例翻訳や統計翻訳は言語の規則の作り方がちょっと違う。
今度は工学者の方が怒るかもしれませんが、だいたいそのように理解していて、そう大きく間違いではありません。
それに対して、概ね正しい翻訳の図式は、もともと誰かが何かを書いたという、歴史社会的な行為です。
その前にすでに関連する文書群があり、その中で、原テクストが書かれる。
翻訳者は何をするか。
ちょうど原言語において翻訳対象となる文書が位置づけられている文書群に対応するような文書群が翻訳言語の中で存在する。しない場合は、後で扱います。
するとして、それに対して、“原テクストが位置づけられる固有テキスト群”と“原テクスト”の関係を、“原テクストが位置づけられる固有テキスト群”に対応する、“対象テキストが位置づけられる固有テキスト群”と“翻訳テクスト”との関係に移し替えるというの

が翻訳という行為です。
そのためには、抽象的な言語の規則ではなく、歴史的に構築された参照テクストやツール文を使う、典型的なのが、さきほどの人種差別撤廃条約の翻訳については、もっとも正しい翻訳の出発点は外務省から正訳を探すこと、ということになります。

・スライド 26
「翻訳とはそもそもどんな行為か（プラトニックに）（２）」
そう考えるとミッシェル・フーコーが言った言葉、
言語の分析が、ある言説の事実に関して問う問いとは常に、
「どのような規則にもとづいて、このような言表が作られたのか？また、したがって、
どのような規則によって他の同じような言表をつくることができるのか？」
であるのに対して、言説の記述が立てるのはまったく別の問いであって、
「このような言表が出現した、しかも、他のいかなる言表もその代わりには出現しなかったのは、どのようなわけなのか？」
という問いなのだ。

こう言われているところの、後者のようなものに「相当するセンスを」というとフーコー研究者がこんどは怒ると思うので、そんなに単純化するといけないと言われますが、そういうセンスを持つことがきわめて大切なポイントとなります。

・スライド 27
「翻訳とはそもそもどんな行為か（プラトニックに）（２）」
ここまでで翻訳とはそもそもどういう行為かが定まってきました。扱うものは言語表現テクストという歴史社会的な産物、これが人間の翻訳で、機械翻訳はそれを剥奪した、言語としてみた言語表現です。
典型的なことは、ここから何が分かるかというと、小学生や中学生に対しても、日本語を母語としない人に対しても機械翻訳は同じ結果を出します。
なんでか？
大雑把な意味で、歴史社会的な文脈を無視しているから。
立場としては、翻訳は先ほど言った翻訳した文書の歴史社会的な位置を定める意思決定をするのが翻訳行為であって、機械翻訳は「言語」の計算である。
翻訳の問題点は、対応する位置を定めるとは言ってもそもそも対応する目安がない場合はどうするか。
これについては、時間があったら後ほど。

・スライド 28
「翻訳とはそもそもどんな行為か（プラトニックに）（２）」
もう１点、翻訳の典型的なパターンを挙げます。
村上龍の「69」です。
原文は、「…わかった、浅丘ルリ子よりきれいに映さんばだめばい」です。これが翻訳の方

では、
「Yeah, that makes sense. But listen, man, you better make her look good. Better than Brigitte Bardot.」になっています。
まさに日本の文脈で浅丘ルリ子に相当するのは、ブリジット・バルドーなんです。と強引にしてしまった。でもこれは、
半分は良くて・・・怒る人もいますが、半分は悪くて。
半分良いところは、このほうが、歴史社会的な現実の文脈の中で英語圏の人に対応関係を位置づけることが出来る。
ところが次でこけます。
松井和子と石原裕次郎はそのままなのです。
とんでもない話でそのままです。
これに類する例です。
多和田葉子さんの詩の中の
「畳」という言葉をドイツ語の翻訳者がそれをドイツ語の「絨毯」と訳しました。誰かが文句を言ったときに、でもこの詩の中で「畳」は少し前の日本の家族のイメージを映し出したかったので、ドイツでも「畳」がはやっているけれど、それをそのまま「たたみ」と訳してしまうと・・・ドイツでは新しいファッショナブルな若いカップルなどが畳を好んで使うので、家族の記憶とか、そういうものを強調するのに、ドイツで「畳」に相当するのは「絨毯」なので、絨毯でいいということを言っていて、それらも翻訳というのがかならずしも言語の計算にとどまるものではなく、社会歴史的な文脈を含むということを示しています。

・スライド 29
「機械翻訳の精度が仮にもっと向上したら」
機械翻訳の精度が仮にもっと向上したらどうか。
今みたいな事を考えると大量のデータを集めて、抽象的に言語の計算をするのではなく、歴史的文脈などもとっておいて、探せるようにしたら機械翻訳の精度は、いわゆる
X 文 Y 訳ではなくて、翻訳そのものとしても向上できるのではないかと思われます。
X 文 Y 訳としての精度向上を機械翻訳研究はやってきて、さらに翻訳として存在するためにも、もうちょっと精度の向上はできるかもしれないと思います。
ところが実はあるところからは、精度の問題ではないことが発生するだろう、と思われます。
どういうことかは、少し説明しますが。
精度がどんどん向上してもダメでなのす。なぜか。
１つめは、よく考えてみると今、私が話している文、これストリーミング配信されますが、聞き直してみるほど奇特な方はいらっしゃらないと思いますが、試しに聞き直してみてください。
そうすると私の話している文は、一応私は日本語の母語話者といわれているにも関わらず、言語学的な日本語の文法規則から考えるとかなり逸脱した表現、或いは「てにをは」を間違えたり言い直したりしています。

それにもかかわらず、何かこんな所に立って話す機会を得て、しかも内緒ですが、謝礼ももらえます、不思議です。
何でか、これは実を言うと人のコミュニケーションはそんなに完璧ではない、という極めて陳腐で凡庸な結論にいきつきます。

もうひとつは、先ほど機械翻訳で、犬に対しても同じ言葉を使うということがありました、私は犬に対して同じ言葉を使いますが、そうすると時々咬まれます・・。
人間の場合は、話し手の自己認識と、その話し手の自己認識に対する相手の理解が、コミュニケーションを成り立たせる前提だというのを、大雑把にいうとグライスさんのコミュニケーションの公準みたいのがあって、私自身はそれに逸脱しまくっているのですが、そういう規則があります。
つまり、誰に対してどんな話をするか、誰に対してどんな話をするということを相手は何となく分かっていて、相手もそれを受け取っていて、さらに、相手がなんとなく分かっていることを話し手も分かっていて、相手が何となく分かっていることを話し手が分かっているとういうことを、相手も分かっていて、さらにそれをこっちも何となく分かっているような関係。これは最低限いつ、どこで、どんな時にどんな言葉を使うかという使い分けと、ある程度対応しています。
そういえば、人間同士の対話も実はそうだということを今、思い切り言いまくってしまいました。
この両者から、あるところからは精度の問題ではないというお話が出てきます。
なぜか。結局、コミュニケーションの原理に関わる問題は、一定水準の言語的な力、
操作性は当然必要ですが、そこから先は、ちょっと別なのではないか。
何でか。
もしそうじゃなかったら、生まれたての赤ん坊が、言葉を出来るようになることは決してありえません、最初は言葉がないわけですから。
言葉がしゃべれないからおまえとは話せないとなったら、永遠に言葉は話せない。

・スライド 30
「お話しの概要」

・スライド 31
「結局は、コミュニケーションの原理に関わる問題」
ちょっと回り道をします。
提起されていたが検討されていない問題。
　“翻訳対象側に翻訳文書を位置づける既往文書群がないとき”　や　“機械翻訳の「自己認識」”
我々が誰かに話すときに、ああこういう人に話ししているんだという認識をどうしてやるかとかいう問題。
それからもう１つ、理屈の問題で、コミュニケーションに関わる問題ですが、文化の違いを翻訳できないというようなことについて少しだけ話して、コミュニケーションの原理に

関わる問題の話を終わらせます。

・スライド 32

「結局は、コミュニケーションの原理に関わる問題」

文化が違えば翻訳は不可能という人がいますが、それは間違っています。
神学論争や勝手な思い込みではなく、それなりに検証し検討することのできる主張としてとらえるならば、間違っていて、なぜならば、検証し検討できる主張なら、「翻訳は不可能です」と言えるためには、可能性としてはある与えられた言語表現に対して無限に存在しうる様々な翻訳について１つ１つ、この翻訳はこのような理由で、正しい翻訳ではないと言えないといけない。

それを顕在的に無限にあり得る様々な翻訳の可能性について、具体的に言うためには、少なくともこの世の中で誰か１人、正しい翻訳を知っていなければいけない。
少なくとも統整的な理念としてはそれが想定されなくてはならない。
だから「翻訳は不可能です」と言ったとたんに、神学的主張でないなら、翻訳が可能であることを前提としていることになります。
もう１つ、通常翻訳は不可能という場合は英語と日本語とか文化的な違いを言うのですが、誰 1 人同じ経験をしたり、同じ言葉を話している人もいません。たぶん皆さんの中で、私と同じような言葉を話す人はいないと思います。
つまり私がこういうことを表現したいと、53 分話してきたような形で表現する人はいない。
つまり文化が違えば翻訳は不可能という主張は、誰であれ人は自分以外とはわかり合えないという主張にたどりつきます。
日本の原風景はお米だと言っても、私は北海道出身だから、田舎の水田ではなくジャガイモ畑です。仮に同じ言語を話していてもそうなります。少なくとも私自身は
文化が違えば「翻訳は不可能です」と主張する日本人より、「翻訳できる」というフランス人やケニア人などのほうがわかりあえるような気になります。これも幻想かもしれません。
ですから翻訳は不可能というのは間違っています。
コミュニケーションの原理に関わる問題の周りをゆれているだけで申し訳ないのですが、

・スライド 33

「結局は、コミュニケーションの原理に関わる問題」

翻訳対象側に翻訳文書を位置付ける既往文書群がない場合。どうするか。個別の対処方法としては、注を付けます。
エリオ・ヴィットリーニというイタリアのネオレアリズモの作家の「シチリアでの会話」という本が数年前に岩波文庫から出ていますが、3 分の 1 くらいが「注（解説）」です。実際にそうしないと歴史社会的文脈にこの作品が位置づけられないからです。
ここからがコミュニケーションの問題で、対応する文書に位置付けることを、常に過去だけ見ていては絶対に問題は解決しません。
でも、全体としてみると未来に向けてゆっくりと時間をかけて慣れるというのがポイントになります。

こういうことを言うときれい事といわれますが、大きなポイントです。どういうことか。
考えてみたら、150 年前には、皆、漢詩が書けて、書道もうまくて、ちょんまげだったりして、150 年経ったらマクドナルドに行くのが普通で、アメリカでもマクドナルドで……
となっています。
ゆっくりと時間をかけて慣れることで、文化というのはそれなりに混ざり合う、交流する。
よく考えたら、単一言語の中でも最初に生まれた赤ん坊は、何も言葉が分からない、言葉ができるようになるため、驚くほど早く言語は習得しますが、ものすごい時間をかけて父親や母親、周りの人が赤ん坊に話しかける、それほどゆっくりではないかもしれないが、最初にコミュニケーションを可能にするためには、あらかじめ条件があり、それが満たされているからコミュニケーションが可能なのではなく、ゆっくりと時間をかけて慣れていくことで、あとで見たらコミュニケーションができていた、と思われるようになる。
こう考えるほうが妥当です。
そうすると、人間に対しては、そういう形、とりわけ同一言語を共有する人たちに対しては、そういうものをある程度受け入れているが機械に対しては、それを受け入れることができない。
機械側の自己認識と、私たちの機械に対する認識の中に、ゆっくりと時間をかけて慣れていくというサイクルが確立していないというのがポイントになっています。

・スライド 34
「お話の概要」

・スライド 35
「翻訳・情報流通・多言語バリアフリー」
次に最後の問題です。
翻訳情報・流通・多言語バリアフリーの話。
まず、以上の自己認識を頭におき、改めてもしもうまく自己認識的なサイクルを、自己認識と自己認識に対しての他人の認識と、自己認識に対しての他人の認識の認識というものをコミュニケーションに参加する人たちがある程度、ゆっくりと時間をかけて慣れる中で、コミュニケーションを前に進めることができるとして、そういう中では、機械翻訳は一部の多言語バリアフリー化に使えます。
“ホテルの情報を自分で確認をする”というのは、
今、私のところに来ているヨルダンの学生は日常的に旅行に行くときは、日アラビックや日仏の機械翻訳を使っています。
他にロシアのオンラインサイトから商品を購入するとき。
これらについて、人間の翻訳者が威張って或いは、威張らずに「機械翻訳はダメ」と言っていても、これらの状況にいる人の手伝いは絶対にしてくれません。
翻訳企業のページを見ると最低発注単語数というのがありますが、ロシア語のウェブサイトで、３単語だけ訳してという発注は受けてくれません。でも機械翻訳なら文句を言わずにやってくれる。
これにも何とか慣れることが出来るはずだと考えましょう。

ちょっとカッコつけた言葉を言いますと
「歓待とは忘却の振舞いではなく、あくまでも差異に対する意識化の振舞い」である。

・スライド 36
「翻訳・情報流通・多言語バリアフリー」
ただし、機械翻訳が制度の問題として単純に自己認識とコミュニケーションのサイクルの問題としてではなく、制度の問題として使えないことがあります。
契約書や条約の翻訳などです。広い範囲の少し複雑なほとんどのテクスト。
人間の翻訳は使えます。ただしトラブルが起きうるという意味では、常に正しいわけではなく、トラブルが起こったときの手続きが、時間の中で確立したという観点からです。
ただし、そのように馴染んだ人間の翻訳が使える範囲はごくわずかです。

・スライド 37
「翻訳・情報流通・多言語バリアフリー」
以前は人間以外が翻訳するなんて思いもよらなかったわけです。そうすると、多言語情報流通は限られた対象、限られた言語、限られた人の中で行われていた。
川口　篤さんが戦後、まもなくしてアンドレ・ジッドの本を岩波文庫から出したら家が建ったということがありました。現代はそんなことはありません。
メガネができる前は視力が低い人はバリアで取り囲まれていたのですが、バリアフリーという観点から、多言語の分野は、機械翻訳やそれに類する翻訳を支援するメカニズムの中で広がっています。

・スライド 38、39
「翻訳・情報流通・多言語バリアフリー」
大雑把な図式をあげますと、
簡単な情報をとるときは機械翻訳でよいでしょう。それに対して、言語表現の難しいところ、文学とかは翻訳者が必要です。その間で人間と機械が翻訳を協力しようという領域はたくさんありえる。これがさきほど言ったコミュニケーションをうまく繋ぐこと、コミュニケーションのサイクルを確立し、ゆっくりと時間をかけてなじんでいくような社会制度をこれから構築していけばということですが、そのような目的で「みんなの翻訳」というのを、とりわけ NGO、NPO 向けにやっています。

・スライド 40
「扱わなかった問題」
扱わなかった問題が最後にいくつか残っています。
そもそも、今、最後のほうで言っていたきれい事はどうやるねん、ということです。
人間と機械の協調と翻訳はどうやるのか。
その他に、最後に付け加えることがあります。
1 つ目、人間をバカにしたシステムと人間を賢くするシステム。

2つ目、消費者、商品、参加者としてみる。
この前に、ここで一方でバリアフリーのところで機械翻訳が使えるといいました。これはかなりきれいごとで、条件が必要です。
まず、バリアフリーとは何か。
欠落を補うためのメカニズム、ではなく、私の研究科にバリアフリー研究センターというのがありますが、そういう人と話していると欠落を補うためのメカニズムではない。
これもきれい事ですが、人間の潜在的な能力を引き出すための環境作りがバリアフリーというところが大きなポイントとなります。
これはこのきれいごと的な言い方は、先ほどのゆっくりと時間をかけて未来に向けて馴染んでいくことということと、丁度ぴったり重なっていきます。

ところが、最後にもう1点。
人間をバカにしたシステムと人間を賢くするシステム。
これは、さっき言ったバリアフリーの2つの見方、欠落があるから、補うんだというのは、人間をバカにしたシステムです。
人間を賢くするシステムは、潜在的な力があるが、環境によって発揮されていない面があるから、それを整える。
それに対応して、それを踏まえて後者のようなシステムや機械翻訳を使いましょうというときに、現状をもう1度見直す。
現状をもう一度見直すと、多言語バリアフリーに使えますとのうのうと言いましたが、それにも関わらず、実は Google を始め様々な機械翻訳システムの最も基本的なところでは、「タダ」なんですね。
なんでか。
広告収入があるからです。
そうはいっても、経済の中で、どうしてそんな事がありうるのか。
私たちの嗜好や好みを広告主が買い取ってくれる。つまり今のところ当たり前に使っていれば、多くの場合に機械翻訳は情報の消費者として多言語情報を消費する。
それを介して自分たちが商品になってしまう。というところがあります。

それをどうやって変えることができるか。一方で、人間の翻訳者は完全に媒介者、生産者というプライドを持っています。
この配置は大きく違うもので、先ほど、勝手に人間と機械が協力しましょうときれいごとを言ったところについては、一体この協力関係を通してどんなふうに人間が参加者になっていけるか。
また、ここで「参加者」という言葉は、コミュニケーションをゆっくりと時間をかけて将来にむけて成立させていくプロセスに自分がどのように関与するかという話になっていきます。

最後にまとめます。
人間と機械の翻訳の違い。

人間の翻訳は翻訳である限りは歴史社会的な言語表現、テクストを扱う。
機械は言語の計算をする。
一番最後のところで、機械翻訳と人間の翻訳の違い。
人間の翻訳における翻訳者、翻訳利用者は、生産者、媒介者である。
それに対して受容者が消費者。
機械翻訳も受容者が消費者なのですが、機械翻訳の場合は、機械翻訳の
産物がただ、ということは、消費者というだけでなく、利用者はさらに商品になる恐れがある。
その中で 2 つのこと、商品になる恐れを避けながら、さらに人間と機械が協力することによって、単なる消費者ではなく、少しだけコミュニケーションに参加していくというルーチンをどうやって作ることができるか、これが一番大きな社会的な課題であるとともに、機械翻訳についての技術を簡単に説明しましたが、Google の翻訳が今、ここまで手足を伸ばしている中、実は技術的な課題にもなっています。
8 分もオーバーしましたが、ありがとうございました。

・・・

スペイン語の代名詞の「lo」とかが、it なのか him なのかでコケる以外は、そこそこ普通の手紙などはうまくいきます。
言語の計算では、英語から日本語、日本語から英語など、構造上大きく違うのはきわめて難しくうまくいきません。
日本語は、朝鮮語／韓国語と構造がきわめて似ているので、非常に精度よく機械翻訳もいきます、かなり意味が通じる程度のものをかなり広い範囲で作れます。
英語から日本語への翻訳と、カタロニア語から日本語への翻訳は、言語の計算の観点からいうと難しさは大してかわりません。
でも人間の翻訳の観点からいうと、英語を日本語に訳するほうが、カタロニア語を日本語に訳するよりはるかに楽なのです。
これはなぜか、歴史的に英語の文献は膨大に日本語に直されてきたが、カタロニア語の文献は直されてきていないので、日本語側で、翻訳の産物となる文章を位置付ける文脈が充分構成されていない。
ですからさきほどの、エリオ・ヴィットリーニの「シチリアでの会話」の 3 分の 1 ぐらいが「注（解説）」に費やされているのはそういう理由からです。
答えになっていますかね。

・質問２
司会／2 つ目の質問。
日本語の学習者向け、外国人が日本語を学ぶときの教材を作る、普通の日本語から
やさしい日本語に訳すことに、関わっていらっしゃる方がいます。
普通の日本語を分かりやすい日本語へ、そういう翻訳を機械でできるか、その可能性に
ついてお聞かせ願えればと思います。

回答：影浦／ある程度、機械による手助けはできると思います。
人間が、言語間の翻訳をする時、機械による手助けするのと多少類似していて、言い換えの例等を大量に集め、それなりに参照できるようにすれば、少しは効率的になる可能性が現在あるかと思います。

司会／その質問にも関係しますが、今日の話のメインになると思いますが、機械翻訳ではなく、人間が翻訳するというのは、歴史のあたりの巻物のエリアを別の巻物のこのエリアにマッピングするというのが、本質であると私は理解しました。
機械が難しいところ、巻物の場所が一致することを、別の表現で言うと、文脈、
コンテクスト、これが歴史というように、今日、説明いただいたことと対応すると
思います。
文脈、コンテクスト、別の表現では、意味、という言葉になるかどうか分かりませんが。
そういうものを機械が抽出する技術、試みは、なされている気がします。
そういうトライ、機械翻訳の試みは、どれくらい進んでいたり、実現しているのでしょうか。

回答：影浦／最終的な条件が微妙なのでうまく言えませんが。
まず文脈という言葉でさされるものがいつもいつも、機械で処理する場合は、残念ながら今のところ歴史や社会から浮いてしまった、人間の側で暗黙の前提とした上で、言語的文脈になってしまいがちです。
ただし、先ほど私が言った文脈は言葉で表せない抽象的なもの、操作できないもの、
今ここにいることの１回性だとまでは言うつもりはありませんが。
そういう観点からいうと、単純なことで進むこともあります。言語処理屋さんは、言語的文脈を追いかけています。
考えてみればトップダウンに、これは子ども向けなどと、言語の属性としてではなく、
テキストの属性としてテキスト分類でやっていることが一部ありますが、それをどんどん、社会の中に位置付けていけばある程度、これまでより柔軟な形で、さきほど機械翻訳では相手によって言葉をかえられないといいましたが、そのようなことにも、対応できる
ルーチンを作ることができる。
更にそれを計算して、自己学習というか外からの情報を入れて学習させれば、「あ、こんな文章はできません」などとプロの翻訳者は言いますが、それと同じようなものを機械も
極端な例については言えるようになり、「これはできません、ごめんなさい」と言えるようになる。
これはものすごく大きな進歩と思います。
とっても大切な面白いことを言おうと思ったんですが、……なんで忘れちゃったんでしょう。あ、アンドレ・ジイドの「狭き門」です。つまんない作品なんですが。大売れに売れた。

文脈の話、最終的には、ちょっとした問題が残っていて、そのような形で、私は出来ますとか出来ませんとか言った時に、人間側が機械と対話する意志があるかが、大きな問題になります。というのは人間と機械の大きな違いは、きれい事的に言った、歴史社会的な文脈ということを１つ言って、更にゆっくりと時間をかけて馴染む。これは過去の歴史社会的な文脈を固定のものとしてその中で私達は理解します、というだけでなく、理解できないことについては、これから理解できる文脈を培っていきましょうということに関わってくるわけですが、
計算機ではこれはできない、使えないと、私たちはやめちゃうところがあるので、そこを辛抱強く付き合うことができるのか。
そしてつきあうことが出来ることに対応できるような計算機側の準備ができているかが大きな問題になってきて、そこは単純に技術の問題だけではなく、計算機側の技術に対しての自己認識の問題と、計算機側の技術に対しての私たちの評価の問題と態度の問題、それから計算機側の技術に対しての自己認識に対する私たちの評価と態度の問題というのが関わってくる。その部分は純粋に技術的問題ではないと思います。

・質問３
司会／続いての質問です。
機械翻訳の難しいところの１つとして、小学生に向けての翻訳なのか中学生に向けての

翻訳なのか、研究者向けなのか、誰に対する翻訳かが全部一律になってしまう。
それが問題でもあり、特徴でもあるということでした。
例えば、Google や Excite の翻訳はある程度、データベースがあります。この文とこの文が対応している。
この日本語表現とこの英語表現が対応するという、もしもこの小学生言語表現とこの中学生言語表現とこの大人向け言語表現が対応しているというデータベース、日本語間だとして、そのようなデータベースを Google や Excite が作ったとしたら、英語から日本語の小学生向け、中学生向け、大人の日本語向けというように相手が誰かという認識をしなくてはならないが、誰が使っているかに応じて、ある程度適切にその人向けの言語に機械が翻訳することが可能と思いますが、それは、機械翻訳の壁を打ち破ることになるのか、それとも歴史というか、まだまだ近づけないものなのか、どういう位置付けと考えるか？

回答：影浦／技術的なポイントとしては実はそれに類するシステムはすでにあって、漱石の文章を鴎外にするかとか、女子高校生向けにするとかはあるのです。それはそこそこうまくいったりしています。
機械翻訳は誰に対しても同じ言葉を話すと言いましたが、同じ言葉をしゃべる言葉の表現のレベルでは、ある程度人に応じて変えることはできると思います。
もう一遍戻って、人に応じて変えていくという機械の振舞いを人間の側がどう捉えるか。
あるいは、機械の側が人によって変えるというように、「小学生向けに訳してね」と言えば、変えてくれるのか、それとも小学生だと思って変えるのか。
さらに、私たちは話をして「こいつ、わかってないな」と思ったら別の言い方をしたりする、そのように機械が自己調整できるという、少し上位のレベルの問題としては、現在の技術とはちょっと違ったところで考える必要があります。
でも、言い換えてくれるシステムは、遊びのシステムとしては、結構いっぱいあって楽しいので今度やってみてください。

司会／ありがとうございます。

・質問４
次の質問です。
翻訳技術の 1 つとして、今日、翻訳で難しい点の原因となるものとして短い単純な文だとかなり簡単に機械翻訳でも正しくなるが、文章が長くなると支離滅裂になる傾向があるという話がありました。
ということは長い文章をなるべく単純な短い文章に分けていくような技術がもしできたとすると、それとコンビネーションして機械翻訳が少しマシになるというような可能性が感じられるのですが、そういう技術、取り組みはどのような状況でしょうか？

回答：影浦／あまりにも長い文を短い文に分けたり、日本語は、ヘッドファイナルと言って、修飾部分は前からかかるものです。
英語は形容詞と名詞の関係では、修飾部分は前からかかるので、日本語と同じ形です。

修飾節の場合は、日本語は前からだが、英語では which などで、後ろからかかる。その形を、まず英語の側で組み換えて、基本的な単位をうまく取り出す研究はあります。
その限りでは、技術的にはかなりうまくいっています。
言葉を変えると、先ほど、ポルトガル語からスペイン語の翻訳はうまくいく、といいましたが、これは大雑把に言うと、深い解析、構造が大きく違わないので、単純に、比較的機械翻訳がうまくいくような例に相当するような単位に普通に解析していても、自動的にその単位のまとまりをとってくれることが多い。
少し技術的な説明としてはインチキですが、結果をわかりやすく説明するためにはそんな感じです。
構造の大きく違う言語についても、まず構造を近づけるような処理を前処理としてする。
それから対応するものを簡単にとることがあります。

司会／わかりました。

・質問５
次の質問です。今日のいろいろな言語、韓国語、カタロニア語、オーストラリア英語などの例がありました。
世界にはイギリス英語やアメリカ英語も、オーストラリア英語、日本人の話すブロークン・イングリッシュもあります。
いろいろな英語が世界中に散らばっています。
この研究所は私の研究室でもそうですが、いろいろな国の人が集まって、ひっちゃかめっちゃかな英語でしゃべる、これからそういう機会が多くなってくると思います。
言語としては同じ言語表現をしゃべっていても、かなりバックグラウンドにひっぱられたジャパニーズ・イングリッシュとか、ブロークン・イングリッシュがグチャグチャなっている場面である程度日本語的な英語をちゃんとしたブリティッシュ英語にするとか、
インド系の英語をアメリカ英語にするとか。
方言になるのかもしれませんが、同じ言語の中でも、世界の中の国際交流を手助けするような英語間の翻訳のようなものは、必要性があるような気がします。
そのようなものが実現可能とか、実は、もうシステムがあるとか、現状はいかがでしょうか？

回答：影浦／大変大切な質問だと思います。
技術的に答える前に、そもそも同一言語間の翻訳がとりわけ一部のグループはネイティブなイングリッシュスピーカー、母語話者で別のグループはそれを第 2 言語で学んだ、という状況で言語間の変換や突き合わせを支援するときは、方向性が生まれて、その技術支援を定義するという形で研究が進んでいます。
英文校正支援システム、自動英文チェックシステムなどです。
ところが、今の質問はその前提となる、英語をしゃべるならネイティブだという、
実は英文構成システムの基準です。
学術論文を書くならば、ネイティブでアカデミックな学術論文が規範だということが英文

校正支援システムの基準なのですが、これをとってしまいフラットにする。
今のところ、例えばオーストラリア英語の話し手が日本人の英語のようなものを話したいとは思わない。
残念ながら・・・双方向的な同一言語内の変換、翻訳イメージで定義されたシステムはありません。
しかし、技術的には十分可能です。
ちなみに、イギリス人とかで、男性で女性にモテたいためにわざとイタリア語の母語話者がしゃべっているような英語の発音で話す人を私は 2〜3 人知っています。
そういう特別な例はありますが、ノンネイティブの言葉には、そのような付加価値…と言えるかわかりませんが、そういうものがない限り、結局ノンネイティブの言語を含んだ同一言語内翻訳システムは校正システムと定義され、研究されています。

・質問６

司会／ありがとうございます。では、最後の質問です。
機械翻訳の難しいところ、限界点は今日何度も出てきたキーワード、歴史的な背景、文化的な背景との対応関係を理解しなければいけないところにあると私は理解しました。
それは人間同士でも難しいと思いますが、もし今後、機械がどんどん発達して、例えば人間と同じように歴史を歩むような機械が出てきたら、私はロボットの研究をしていますが、人間型ロボットが人間と同じような日常生活をして、同じ生活を何年も、10 年、100 年と鉄腕アトムのようなものが生活をして、鉄腕アトムのような人間とそっくりな、人間と一緒に歴史を刻んだ機械は、人間と同じような翻訳ができる可能性がありますか？
ちょっと夢物語のような質問ですが。

回答：影浦／おそらく今の質問は、条件の中にすでに答えが含まれているので、そのような場合は、その条件が満たされれば「ある」と答えざるをえないと思います。
きちんとした答えになっているかわかりませんが、質問から外れた形になるかもしれませんが、今でもコンピュータと対話をする。
コンピューターをあたかも人間のように扱い、コンピューター上の登場人物と“だけ”ではないですが、話をして幸せという人はいます。
さきほど言った、機械に対しての自己認識とそれに対しての人間の側の認識というサイクルは、最終的にはそのような形のコミュニケーションが成り立つように我々の意識を変えることになるわけですが、
ただ、現状を見ていると、ちょっとした逆説が観察されます。
あまり強調しても良くないですが、コンピュータと楽しく話すという人は、人間と話すのが苦手だと、それはなぜか。
人間は思いがけない行動をして怖い、ということです。
ここに微妙な違いがあり、ゆっくりと時間をかけて馴染むというのは、最初から自分と同じ状態で自分とわかり合えているという状況ではない。
今のところコンピュータと話ができるというのは逆説的ですが、コンピュータが基本的には壊れない限り、自分の想像通りに振舞って、

あるいは自分がコンピュータをそれなりの自分の想像通りに振舞わせることが出来て、そして想像どおり振舞わせたものに対して、自分の想像を重ね合わせてという形になっていく。
基本的には自分自身の投影である、と言ってしまうと言い過ぎかもしれませんが。
そういう状態であるからこそ、コンピュータと話す。
ところが人間が本当に人と話すときは、そうじゃない。
3 割、5 割、8 割かも知れませんが、そのぐらいは、それなりに話せるが、ちょっと別の思いがけない行動や、ちょっと別のああ驚いた、ちょっと別の考えたことないほど面白い、ちょっとすごい嫌、そういうこちらの予想が付かない状況があるからこそ、人間同士のコミュニケーションだ、という妙なところがあります。
だからこそ、完全に固定的な過去の参照軸で同じことを話すのではなく、未来に向けて話していく価値があるわけですし、さらに……何を言おうとしたのか忘れました。
ですから鉄腕アトムや何かについて、それを含めてというのが、稲邑先生の条件付けだったと思います。
もしそうなら、コミュニケーションにおいては人間と同様に人間側も機械側もみなすことができる中で、異なる言語を話す機械、通訳してくれる機械を普通に受け入れられると思いますが、ただ、この議論はいささかトートロジーであるかのように思います。……。

司会／尽きない話題ですが、残念ながら時間をオーバーしてしまいました。
このへんで質疑応答を終わります。
影浦先生、ありがとうございました。

（拍手）