Typology of Paraphrases and Approaches to Compute Them

Atsushi FUJITA
Future University Hakodate, JAPAN
http://paraphrasing.org/~fujita/
What’s paraphrase?

- Intentional definition
  - e.g., LDOCE

\begin{enumerate}
\item \textbf{(v)} to express in a shorter, clearer, or \textit{different way} what someone has said or written
\item \textbf{(n)} a statement that expresses in a shorter, clearer, or \textit{different way} what someone has said or written
\end{enumerate}
What’s paraphrase?

- Extensional definition
  - lexical, phrasal, sentential, discourse-level, ...
  - covered all? well-organized?

  **The riddle is solved by me.**  **Employment showed a sharp decrease.**
  **I solved the riddle.**  **Employment decreased sharply.**
  **Emma burst into tears and he tried to comfort her.**
  **Emma cried, and he tried to console her.**

- Scope / boundary
  - Not precisely defined

  **I want some fresh air.**  **My son eats eggplants.**
  **Could you open the window?**  **My son likes eggplants.**
Typology of paraphrases

- Axes
  - Structure
  - Required knowledge
  - Application
  - Sameness and difference of meaning

- Guidepost
  - To clarify how human beings process paraphrases
  - To automate paraphrases (steadily)
    - Clarify required resources for each type
    - Modularize each type for selective use
  - Artificial, so not be crazy
Goal of this talk

- A survey
  - Share the idea
  - Discuss the way of creating typology
    - e.g., Axes
  - Involve people for creating typologies
    - e.g., http://paraphrasing.org/paraphrase.html
Outline

1. Sameness of meaning
2. Linguistically-motivated typology
3. Paraphrases in apps
4. Computation
5. Future directions
Meaning of linguistic expressions

- Semantics
  - Formal semantics
  - Situation semantics
    - Discourse representation theory [Kamp, 81]
    - Mental-space theory [Fauconnier, 85]
  - Lexical semantics
    - Frame semantics [Fillmore, 76]
    - Lexical Conceptual Structure [Jackendoff, 90]
    - Generative Lexicon [Pustejovsky, 95]
Paraphrase in semantics

- A good subject
  - To think of equality
  - Toward semantic computing
    - How to drive semantic frameworks

- Levels of sameness [Sato, 99]
  - Pragmatic meaning
  - Referential meaning
  - Denotation
Sameness of pragmatic meaning

- Illocutionary / perlocutionary acts
  - I want some fresh air.
  - Could you open the window?

Hearer’s interpretation
Speaker wants me to open the window to get fresh air.

- Various interpretation
  - But, only the speaker knows truth
Sameness of referential meaning

- Coreference

Barça’s #10 scored no goal in the last El Clásico.

Lionel Messi scored no goal in the last match against Real Madrid.

in 2008-2011

Barça’s eye view

- May not true in the other situation
  - e.g., Ronaldinho, Riquelme, Rivaldo, ...
  - e.g., against Barça, between Barça and Real

- Discourse-level
  - incl. exophora
  - Cognitive meaning [Milićević, 07]
Sameness of denotation

- Truth-value semantics
  - Tom bought a car from John.
  - John sold a car to Tom.

- Can be carried out
  - Without referring to the communicative situation
  - With linguistic knowledge
  - (With world knowledge)

- Have different connotation
  - Theme / Rheme
  - Formality
  - Emotion (attitude)

[Edmonds, 99][Inkpen+, 06]
Remind the definition

- It supposes some differences

  (v) to express in a shorter, clearer, or different way what someone has said or written

  (n) a statement that expresses in a shorter, clearer, or different way what someone has said or written

- Not exactly same meaning (synonym) [Clark, 92]
- But near-synonym [Edmonds, 99]
Comparison of “blunder” and “error”

[Edmonds, 99]
Aim of making difference

- What’s changed?
  - complex → simple
  - verbose → clear
  - marked → unmarked
  - emotional → neutral

- Reasons why we paraphrase
  - To facilitate communication
    - For confirmation
    - For accelerating understanding
  - To strengthen the solidarity in a community
Paraphrase

- Linguistic variability in conveying a meaning

<table>
<thead>
<tr>
<th>Meaning</th>
<th>Linguistic exp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambiguity</td>
<td>Mouse</td>
</tr>
<tr>
<td>Variability</td>
<td>risk of receiving a severe wound</td>
</tr>
<tr>
<td></td>
<td>possibility to be seriously injured</td>
</tr>
</tbody>
</table>
Mouton & Co. is the publisher that published Noam Chomsky’s *Syntactic Structures* in 1957.

The author of *Syntactic Structures* is Noam Chomsky.

Mouton & Co. gained much with Chomsky’s *Syntactic Structures*.
Entailment

- Not necessarily same meaning
  - \( X \rightarrow Y \)

The author of *Syntactic Structures* is Noam Chomsky.

- e.g., lexical entailment in WordNet [Miller+, 85]
  - march → walk
  - forget → know
  - has started → started
  - Troponymy
  - Backward presupposition
  - Temporal
Inference

- Not ensure even truth

Mouton & Co. is the publisher that published Noam Chomsky’s *Syntactic Structures* in 1957.

Mouton & Co. gained much with Chomsky’s *Syntactic Structures*.

- But useful in some situations [Pantel+, 07]

My son eats eggplants.

Everything is imported to Japan.

My son likes eggplants.

Everything is eaten in Japan.
Summary

- Levels of sameness [Sato, 99]
  - Pragmatic meaning
  - Referential meaning
  - Denotation

- Related concepts
  - Entailment: paraphrase $\iff$ bi-directional entailment
  - Inference: entailment $\supset$ always-true inference
1. Sameness of meaning
2. Linguistically-motivated typology
3. Paraphrases in apps
4. Computation
5. Future directions
Rough classification

- Names used in papers
  - Lexical / Phrasal
  - Syntactic
  - Sentential

- Classification in [IWP, 2005]
  - Phrase-level
  - Sentence-level
  - Discourse-level

Not necessarily atomic, because methods and results are centered
Our linguistically-motivated typology

- Focused on denotation
  - Explainable referring to
    - The given context
    - Linguistic knowledge
  - Ignored differences in connotation

- 5 types based on
  - Influenced scope
  - Generality (or productivity)
Småland, which is located to the south-west of Stockholm, is called “The Kingdom of Glass”. The reason is that there are sixteen glass manufacturers in this area.

Småland is located to the south-west of Stockholm. It is called “The Kingdom of Glass”. The reason is that there are sixteen glass manufacturers in this area.

Note down the number. Otherwise, you may forget it.

Note down the number. If not, you may forget it.
[B] Extra-clausal paraphrase

- Cleft → non-cleft
  - It was his best suit that John wore to the dance last night.
  - John wore his best suit to the dance last night.

- Head-switch (clausal complement ↔ modifier)
  - The conference venue is the building whose roof is red.
  - The conference venue is the building with red roof.

- Move of negation
  - Your application is canceled if you do not reply.
  - Your application is not canceled if you reply.

- Embedded ↔ coordinate, reordering, etc.
[C, D, E] Intra-clausal paraphrases

Generalizable

X solve Y \rightarrow Y is solved by X
X gives Y a fright \rightarrow Y is frightened of X
X is in our favor \rightarrow X is favorable to us
X show a sharp decrease \rightarrow X decrease sharply
X be the author of Y \rightarrow X wrote Y
X comfort Y \rightarrow X console Y

Non-generalizable

pass away \rightarrow die
burst into tears \rightarrow cried
[C] Pure syntactic paraphrase

- **Inversion**
  - If I had money enough, ...
  - Had I money enough, ...

- **Move of adverb**
  - She can speak English fluently.
  - She can fluently speak English.

- **Paraphrase of negation**
  - He drank nothing but famous spirits.
  - All he drank were famous spirits.

- **Less variation**

Independent of the succeeding clause
There’s a risk of receiving a severe wound.
There’s a possibility of receiving serious injury.
Emma burst into tears and he tried to comfort her.
Emma cried, and he tried to console her.

Real Sociedad snapped a two-game losing streak.
Real Sociedad got points for the first time in three games.
[D] **Morpho-syntactic paraphrase**

- Seems to be syntactic paraphrase
  - But have lexical constraints to some degree
    - John smeared **paint on the wall**.
    - John smeared **the wall with paint**.
    - Employment **showed a decrease**.
    - Employment **decreased**.

- Required information
  - Lexico-semantic information
    - Fine-grained argument structure
    - Lexical derivation, antonym, etc.
  - Selectional preference, collocation
### Kinds of [D]: Verb alternation

- **Passive to active**
  - The riddle **is solved by** him.
  - He **solved** the riddle.

- **Dative alt.**
  - Bill sold a car **to** Tom.
  - Bill sold Tom a car.

- **Locative alt.**
  - John smeared **paint on** the wall.
  - John smeared the wall with **paint**.

- **Source alt.**
  - The well gushed **oil**.
  - Oil gushed from **the well**.

- **Reciprocal alt.**
  - The car collided with the bicycle.
  - The car and the bicycle **collided**.

- **Transitivity alt.**
  - Janet broke the cup.
  - The cup **broke**.

[Levin, 93]
Kinds of [D]: Category shift

- Light-verb construction (N ⇔ V), A ⇔ Adv
  - Employment showed a sharp decrease.
  - Employment decreased sharply.

- Adj ⇔ V
  - I visited a priest in the old temple.
  - I visited a priest in the olden(ed) temple.

- Adj ⇔ N
  - I feel drowsy.
  - I have a drowsiness.
Kinds of [D]: Structural alternation

- **Head-switch (NP), N ↔ V**
  - We need an *improvement* of recycling *system*.
  - We need an *improved* recycling *system*.

- **Head-switch (VP), V ↔ Adv, N ↔ V**
  - He *hurried* to check it.
  - He *checked* it in a *hurry*.

- **Move of quantifier**
  - We performed *two* transactions in this morning.
  - We performed transactions *twice* in this morning.
Summary

- A linguistically motivated typology
  - [A] Extra-sentential
  - [B] Extra- Clausal
  - [C] Pure syntactic
  - [D] Morpho-syntactic paraphrase
  - [E] Lexical (word, phrasal)

- Focused on denotation
  - Atomicity
  - Scope
  - Generality
Discussion and issues

- On the typology
  - Less [C] Pure syntactic paraphrases
    - After all, inter-clausal vs intra-clausal (within a VP)
  - Treatment of indecomposable ones

- Lexical semantics for [D]
  - FrameNet [Baker+, 98]
  - VerbNet [Kipper+, 00]
  - Lexical Conceptual Structure [Jackendoff, 91]
  - Generative Lexicon [Pustejovsky, 95]
Outline

1. Sameness of meaning
2. Linguistically-motivated typology
3. Paraphrases in apps
4. Computation
5. Future directions
Map of apps

Paraphrase Generation

- Writing aid
- Reading aid
- Post-process for MT

Pre-process for TTS

Summarization

- Multi-document summarization

Paraphrase Recognition

Consumed by human

- Look up TM
- QA
- IR
- IE
- DM

Consumed by machine

Pre-process for MT

inside of MT
Requirements for application

- Target types of paraphrases
- Differences accepted
  - Connotation
    - Theme/Rheme
    - Formality
    - Emotion (attitude)
  - Denotation
    - Entailment
    - Inference
- Full-auto / consumed by human
Computing paraphrases for machine

Multi-document summarization [Barzilay, 03]
Computing paraphrases for machine

- Pre-edit for machine translation [Shirai+, 98]

- Data is a plan that distributes freely.

- We plan to distribute the data freely.

- Not only paraphrase, but also anaphora resolution

- Entailment / inference cannot be not applied
Computing paraphrases for machine

- Data mining
  - Summary of events [Izumi+, 10]
    - Light-verb construction
    - Keep factuality, but not some aspectual info.
      - try to get the first prize ≠ get the first prize ≠ began to repair ≠ repaired = has started = started
  - Collecting instances of plausible events
    - Discover unknown unknowns [Torisawa+, 08]
    - Build statement maps [Murakami+, 09]
Computing paraphrases for human

- Writing aid (information dispatching aid)
  - Showing alternatives [Max+, 08]
    - Easier, clearer, more-decorative, etc.
  - Automatic rewrite
    - Normalization of specific documents
      - e.g., technical manuals, health reports

- Reading aid (information consuming aid)
  - Simplifying texts [Carroll+, 98][Canning+, 99][Inui+, 03]
  - Adding explanatory information
    - e.g., gloss of words, related terms
区長のあいさつ

平成15年1月第1回議会定例会区長挨拶総言

平成14年（2002年）2月19日

平成14年の区議会の開局に当たりまして、政治運営の基本的な考え方についての
説明、区民の理解に協力を心からお願い申し上げます。

区長にあるべき役割は、区民の生活向上に向けた活動を主導し、政治運営の
機能を十分に発揮することです。区議会の活動を通じて、区民の要求に応え、
区民の声を反映した区議会の活動を心からお願い申し上げます。

区長は、区民と役員との関係を良好に保ち、区の運営を円滑に進め、区民の
生活向上に努めていきます。また、区民の声を反映した区議会の活動を心から
お願い申し上げます。

平成15年1月第1回議会定例会における区長の挨拶のまとめ

平成14年（2002年）2月19日

平成15年1月第1回議会定例会の閉会に当たりまして、区長のあいさつについて、
区民の声を反映した区議会の活動を心からお願い申し上げます。

区長は、区民の生活向上に向けた活動を主導し、政治運営の機能を十分に発揮
することです。区議会の活動を通じて、区民の要求に応え、区民の声を反映した
区議会の活動を心からお願い申し上げます。

区長は、区民と役員との関係を良好に保ち、区の運営を円滑に進め、区民の
生活向上に努めていきます。また、区民の声を反映した区議会の活動を心から
お願い申し上げます。

平成15年1月第1回議会定例会における区長の挨拶のまとめ

平成14年（2002年）2月19日

平成15年1月第1回議会定例会の閉会に当たりまして、区長の挨拶について、
区民の声を反映した区議会の活動を心からお願い申し上げます。

区長は、区民の生活向上に向けた活動を主導し、政治運営の機能を十分に発揮
することです。区議会の活動を通じて、区民の要求に応え、区民の声を反映した
区議会の活動を心からお願い申し上げます。

区長は、区民と役員との関係を良好に保ち、区の運営を円滑に進め、区民の
生活向上に努めていきます。また、区民の声を反映した区議会の活動を心から
お願い申し上げます。
## Paraphrases in apps

- Typology and modularization are necessary

<table>
<thead>
<tr>
<th></th>
<th>IR</th>
<th>IE</th>
<th>DM</th>
<th>MT</th>
<th>Writing</th>
<th>Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>[A] Extra-sentential</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[B] Extra-clusual</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[C] Pure syntactic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[D] Morpho-syntactic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[E] Lexical</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Focus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entailment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Outline

1. Sameness of meaning
2. Linguistically-motivated typology
3. Paraphrases in apps
4. Computation
5. Future directions
Toward full-automation of paraphrasing

Phase 1. Knowledge development
- Handcrafting patterns
- Automatic acquisition (corpus, Web)

Phase 2. Use of knowledge
- Segmentation and disambiguation
- Applicability check in the given context
  - Grammaticality
  - Semantic appropriateness
  - Equivalency of meaning

Phase 3. Tuning for apps
- e.g., simplification, reduction of homonyms, etc.
Paraphrase Acquisition

1st phase toward automatic paraphrasing
Previous work

- Handcrafting patterns
  - Transformation rules [Mel’cuk+, 87][Dras, 99][Jacquemin, 99]
  - Thesaurus (of words) [A lot of work]

- Automatic acquisition
  - Distributional similarity in a single corpus
    [Lin+, 01][Torisawa, 01][Hagiwara+, 06], etc.
  - Alignment of parallel/comparable/bilingual corpus
    [Barzilay+, 01][Shinyama+, 02][Pang+, 03][Ibrahim+, 03][Dolan+, 04]
    [Bannard+, 05], etc.
  - From the Web [Szpektor+, 04]

- Implicit modeling
  - Statistical translation model [Quirk+, 04][Bannard+, 05]
  - Tree kernel [Collins+, 01][Takahashi, 05]
Handcraft rules/patterns

- For a sentence
  - Transformation grammar [Harris, 81]
    
    NP1 V1 (+AUX) V2 (-AUX) NP2
    \[\rightarrow\] NP2 V1 BE V2-PP by NP1

  - Meaning-text Theory [Mel’čuk+, 87]
    
    \[\text{I} \leftrightarrow \text{II}\]

    \[\text{Oper}_1(S_0(X))\]

    \[\text{VP} \leftrightarrow \text{Light-verb construction}\]

- Various types of rules [Takahashi+, 01]
Extract from thesaurus

- Near-synonyms: words within the same synset
  - e.g., WordNet [Miller+, 85]
    - 02526085: achieve, accomplish, attain, reach
    - 05793554: basis, base, cornerstone, foundation, ...
    - achieve ⇔ accomplish, base ⇔ basis
  - Just near-synonym [Clark, 92]
    - Subtle difference [Edmonds, 99]
    - Static synonymy apart from context [Fujita+, 00]
  - How to enlarge thesaurus?
    - Neologisms google (v) ⇔ search Web using Google
    - Named entities Future University Hakodate ⇔ FUN
Extract from single corpus

- Distributional hypothesis [Harris, 64]
  - Semantically similar words tend to appear in similar contexts.
  - e.g., VP ↔ NP [Lin+, 01][Torisawa, 02]

Compute similarity

- find
- a solution
- mod
- to
- pcomp
- subj
- obj

- solve
- obj

- commission
- committee
- government
- he
- I
- ...

- problem
- crisis
- woe
- crime
- ...

X find a solution to Y ↔ X solve Y
With multiple-sequence alignment

- Multiple verbalizations of proofs [Barzilay+, 03]
- Multiple translations [Pang+, 03]
Extract from comparable corpus

- News articles reporting the same event
  - Named entities as anchor [Shinyama+, 02]

In Hong Kong, the government has announced two more people have died in two more death reported in Hong Kong.
what is more, the relevant cost dynamic is completely under control

im übrigen ist die diesbezügliche kostenentwicklung völlig unter kontrolle

wir sind es den steuerzahlern schuldig die kosten unter kontrolle zu haben

we owe it to the taxpayers to keep the costs in check

under control ⇔ in check
Knowledge for Intra-clausual paraphrases

Generalizable

X solve Y  →  Y is solved by X
X gives Y a fright  →  Y is frightened of X
X is in our favor  →  X is favorable to us
X show a sharp decrease  →  X decrease sharply
X be the author of Y  →  X wrote Y
X comfort Y  →  X console Y

Non-generalizable

pass away  →  die
burst into tears  →  cried

Generate & Validate
Collect

[C-E]
Generate morpho-syntactic paraphrases

- Generation of knowledge [Fujita+, 07;08]
  - Syntactic transformation + Lexical derivation

- \( X \) solve \( Y \) \( \Rightarrow \) \( X \ vig Y \)
  - \( Y \) is solved by \( X \) 
  - \( Y \) be \( V \)-PP by \( X \)
  - \( X \) give \( Y \) a fright
  - \( Y \) is frightened of \( X \)
  - \( X \) give \( Y \) a \( Z \)
  - \( Y \) be \( V(Z) \)-PP of \( X \)
  - \( X \) be in \( Z \)’s \( Y \)
  - \( X \) be \( adj(Y) \) to \( Z \)
  - \( X \) show a sharp decrease
  - \( X \) decrease sharply
  - \( X \) show a \( A \) \( Y \)
  - \( X \) \( v(Y) \) \( adv(A) \)
## Issues and current status

### Issues
- How to cover various types of paraphrases?
  - e.g., knock off each type (typology-based)

### Current status

<table>
<thead>
<tr>
<th>Type</th>
<th>Handcraft</th>
<th>Corpus</th>
<th>Combi</th>
</tr>
</thead>
<tbody>
<tr>
<td>[A] Extra-sentential</td>
<td>○</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>[B] Extra-clusual</td>
<td>○</td>
<td>△</td>
<td>Promising</td>
</tr>
<tr>
<td>[C] Pure syntactic</td>
<td>○</td>
<td>△</td>
<td></td>
</tr>
<tr>
<td>[D] Morpho-syntactic</td>
<td>△</td>
<td>△</td>
<td>Low coverage</td>
</tr>
<tr>
<td>[E] Lexical</td>
<td>—</td>
<td>○</td>
<td>Promising</td>
</tr>
</tbody>
</table>

- Manageable
- Too noisy
- Promising
Toward full-automation of paraphrasing

Phase 1. Knowledge development
- Handcrafting patterns
- Automatic acquisition (corpus, Web)

Phase 2. Use of knowledge
- Segmentation and disambiguation
- Applicability check in the given context
  - Grammaticality
  - Semantic appropriateness
  - Equivalency of meaning

Phase 3. Tuning for apps
- e.g., simplification, reduction of homonyms, etc.
Two aspects of paraphrasing

- Paraphrase recognition/identification
  - Given pair of linguistic expressions $\rightarrow$ label $\in \{=, \neq\}$
    - Theme of machine learning research
      - $\langle\text{give an advice}, \text{advise}\rangle$ $\rightarrow$ $=$
      - $\langle\text{give a copy}, \text{copy}\rangle$ $\rightarrow$ $\neq$
      - $\langle\text{make a copy}, \text{copy}\rangle$ $\rightarrow$ $=$

- Paraphrase generation
  - Numerous outputs
    - $\text{investigate why there was a fire}$
    - $\text{investigate the cause of a fire}$
    - $\text{investigate what started a fire}$
    - $\text{make an investigation into the cause of a fire}$
  - incl. unseen expressions
Paraphrase Generation

Example of 2nd phase toward automatic paraphrasing
Generation includes recognition

Step 1. Candidate generation

Step 2. Assessment

Paraphrase Knowledge

Statistical Models

Rules
Step 1. Candidate generation

- Transfer
  - Approach to MT in '70～'80
    - Assume compositionality
    - Substitute parts of input structure

- Transducer
  - Accept sequence (structure is encoded)
All he drank were famous spirits.

He drank nothing but famous spirits.
Transfer

- At the (shallow) syntax level
  - Minimal standard for various apps
  - Backed up by matured parsing technology
  - Many acquisition methods work at the same level

Discussion

- How wide range can be realized at this level?
- How semantic constraints are incorporated?
  - e.g., lexical semantics for [D]
  - Leave until the assessment step?
Transfer guided by lexical semantics [Fujita+, 04]

BECOME
BE WITH

Semantic transfer

Ken receives an inspiration from the film.

Syntactic transfer

Ken is inspired by the film.
Equivalence explained by lexical semantics

- Recovering meaning using GL framework
  - Computing metonymy and default

\[
\text{John began the book.}
\]
\[
\text{John began reading the book.}
\]

\[
\begin{align*}
\text{ARGSTR} & = \\
\text{ARG1} & = x: \text{info} \\
\text{ARG2} & = y: \text{physobj} \\
\text{QUALIA} & = \\
\text{info} \cdot \text{physobj} & = \\
\text{FORMAL} & = \text{hold}(y, x) \\
\text{TELIC} & = \text{read}(e_1, w, x,y) \\
\text{AGENT} & = \text{write}(e_2, z, x,y)
\end{align*}
\]
Step 2. Assessment

- Because knowledge is static
  - Grammaticality
  - Semantic appropriateness
  - Equivalency of meanings in the context

- Filtering, correction, ranking
  - Rule-based
  - Statistical approach
All he drank were famous spirits.
He drank nothing but famous spirits.
**Statistical assessment**

- **Grammaticality: statistical language model**
  - Collocation
    - e.g., $<V, \text{Slot}, N>$ [Fujita+, 04][Pantel+, 07]
    - Global grammaticality of sentences [Wan, 05]

- **Semantic appropriateness**
  - Compare gloss and context [Okamoto+, 03]

- **Equivalency of meanings in the context**
  - Suitability for the given context [Pantel+, 07][Szpektor+, 08]
Transducer

- Decoding from lattice
  - Multiple-sequence alignment [Barzilay+, 03]
    - Learn whole sentence
  - Statistical machine translation [Quirk+, 04]
    - Use learned phrase table

```
Begin
  detroit
    *e*
    ‘s
    building
  in
  detroit
    *e*
End
  flattened
  levelled
  blasted
  razed
  leveled
  reduced
  to
  ashes
  ground
  to
  ashes
  into
  down
  rubble
```
Issues and current status

- Application of knowledge to a certain context
  - Influence of paraphrase to the context
  - How to deal with generality and idiosyncrasy?

- Two approaches
  - Transfer + assessment
  - Transducer

- Viewpoints of assessment
  - Grammaticality \(\text{Discussed}\)
  - Semantic appropriateness
  - Equivalency of meanings in context \(\text{Not yet explored}\)
Outline

1. Sameness of meaning
2. Linguistically-motivated typology
3. Paraphrases in apps
4. Computation
5. Future directions
Future directions (technical points of view)

Phase 1. Knowledge development
  o How to cover various types of paraphrases?
    ✤ → Not enough
      ✤ Need a formalism and a resource repository

Phase 2. Use of knowledge
  o How to deal with generality and idiosyncrasy?
    ✤ → Some levels on grammaticality
    ✤ → More studies on “paraphrase in context”
      ✤ We ask users in generation-type apps

Phase 3. Tuning for apps
  o How to selectively use each type of paraphrases?
    ✤ → No cross-application platform. Modularization!!
Future directions (linguistic points of view)

- Establishing the way to compile the typology
  - incl. infrastructure: community, portal

- Parallelism
  - 散歩に出かける → 散歩する
  - take a walk → walk
  - faire une promenade → promener
  - dar un paseo → pasear
Thank you

Acknowledgment
My ex-supervisor: Prof. Kentaro INUI
My ex-boss: Prof. Satoshi SATO
http://paraphrasing.org/~fujita/