< IJCNLP 2005, Oct. 11th, 2005 >

Exploiting Lexical Conceptual Structure for Paraphrase Generation

Atsushi FUJITA[1], Kentaro INUI[2], Yuji MATSUMOTO[2]
[1] Kyoto University
[2] Nara Institute of Science and Technology

Generating & Recognizing Paraphrases

- Paraphrases
  - "Alternative ways to convey the same information" (IWP)
- Middleware for a wide range of application
  - Generation
    - Text simplification [Carroll et al., 1999] [Inui et al., 2003]
    - Pre- and post-editing for MT [Shirai et al., 1995]
  - Recognition
    - QA [Hermjakob et al., 2002] [Takahashi et al., 2004]
    - Multi-document summarization [Barzilay et al., 2003]

Variety

- Lexical paraphrase
  - Emma burst into tears and he tried to comfort her.
  - Emma cried and he tried to console her.

  [Barzilay et al., 2001]
- Syntactic paraphrase
  - It was his best suit that John wore to the dance last night.
  - John wore his best suit to the dance last night.

  [Dras, 1999]
- Lexically compositional paraphrase
  - Steven made an attempt to stop playing Hearts.
  - Steven attempted to stop playing Hearts.

  [Dras, 1999]

Lexically compositional paraphrases (in Japanese)

- Paraphrasing of light-verb constructions (LVCs)
  - ***film***-NOM ***him***-DAT ***impression***-ACC ***to give***-ACTIVE
    (*The film made an impression on him.*)
  - ***film***-NOM ***him***-ACC ***to be impressed***-CAUSATIVE
    (*The film impressed him.*)
- Locative alteration
  - ***he***-NOM ***wall***-DAT ***paint***-ACC ***to splay***
    (*He splayed paint on the wall.*)
  - ***he***-NOM ***wall***-ACC ***paint***-with ***to splay***
    (*He splayed the wall with paint.*)

  - syntactically regular
  - semantically compositional
- Category shifting
  - ***room***-NOM ***already***-ADV ***to warm***-Verb-Passive-Perfective
    (*The room has already been warmed up.*)
  - ***room***-NOM ***already***-ADV ***be warm***-Adjective-Present
    (*The room is already warm.*)

Issues

- Issue: to explore…
  - what sorts of lexical properties affect
  - how existing framework of lexical semantics can be used to represent them
- Our attempt
  - Exploit Lexical Conceptual Structure (LCS) [Jackendoff, 1990]
    - Examine current theory and implementation of LCS for Japanese [Kageyama, 1996] [Takeuchi et al., 2002]
    - Develop an LCS-based paraphrase generation model
    - Case study on paraphrasing of LVCs in Japanese

Paraphrasing of LVCs

- LVCs —paraphrase→ single verb phrases
  - Syntactic and semantic properties of two verbs interact

head of semantics　head of syntax

- ***film***-NOM ***him***-DAT ***impression***-ACC ***to give***-ACTIVE
  (*The film made an impression on him.*)
- ***film***-NOM ***him***-ACC ***to be impressed***-CAUSATIVE
  (*The film impressed him.*)

- ***price***-NOM ***exchange***-DAT ***influence***-ACC ***to give***-ACTIVE
  (*The stock price gives an influence to the foreign exchanges.*)
- ***price***-NOM ***exchange***-DAT ***to influence***-ACTIVE
  (*The stock price influences the foreign exchanges.*)

Contents

1. Issues and goals
2. LCS
3. Paraphrasing of LVCs in Japanese
4. LCS-based paraphrase generation model
5. Experiments
6. Conclusion

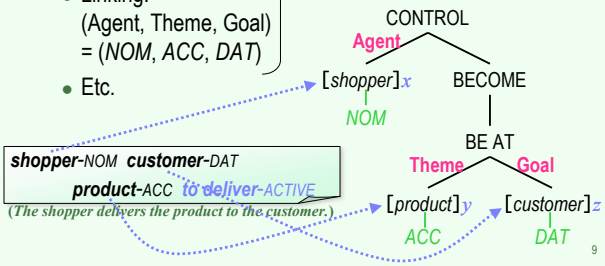
Overview of LCS

- What's LCS? [Jackendoff, 1990]
  - A verb classification which reflects several syntactic and semantic properties of verbs
    - Agentivity: "位置する" (to locate) : Non-agentive
      "遊ぶ" (to play) : Agentive
    - Focus of statement: "与える" (to give) : Agent
      "受ける" (to receive) : Goal
    - Link between syntax and semantics (Linking):
      "遷移する" (to transit) : (*NOM*, *ACC*) = (Theme, Goal)
      "届ける" (to deliver) :
      (*NOM*, *ACC*, *DAT*) = (Agent, Theme, Goal)

Example

- "届ける" (to deliver), "訳す" (to translate)
  - Agentivity: Agentive
  - Focus: Agent
  - Linking:
    (Agent, Theme, Goal)
    = (*NOM*, *ACC*, *DAT*)
  - Etc.

[*y* CONTROL [BECOME [*y* BE AT *z*]]]
[Takeuchi et al., 2002]



Electronic resource and application

- English LCS Verb Lexicon
  - 4,163 verbs / 468 LCS types
  - MT [Dorr, 1997] [Habash et al., 2003], NLG [Traum et al., 2000]
- Takeuchi's Japanese LCS dictionary
  - 1,165 verbs / 16 LCS types
  - Compound noun analysis [Takeuchi et al., 2002]
- Further projects are running (for Japanese)
  [Kato et al., 2005] [Takeuchi et al., 2005]

Contents

1. Issues and goals
2. LCS
3. Paraphrasing of LVCs in Japanese
4. LCS-based paraphrase generation model
5. Experiments
6. Conclusion

Intriguing points

DAT is not necessarily changed to ACC ***"give" != CAUSATIVE***

film-NOM ***him***-DAT *impression*-ACC *to give*-ACTIVE
(The film made an impression on him.)
film-NOM ***him***-ACC *to be impressed*-CAUSATIVE
(The film impressed him.)

TV-NOM ***kids***-DAT *stimulation*-ACC *to give*-ACTIVE
(TV gives kids stimulation.)
TV-NOM ***kids***-ACC *to stimulate*-ACTIVE
(TV stimulates kids.)

price-NOM ***exchange***-DAT *influence*-ACC *to give*-ACTIVE
(The stock price gives an influence to the foreign exchanges.)
price-NOM ***exchange***-DAT *to influence*-ACTIVE
(The stock price influences the foreign exchanges.)

How LCS is expected useful?

- To determine voice and syntactic cases

film-NOM **him**-ACC **to be impressed**-CAUSATIVE

Syntactic cases — Voice

- Voice: how the event is described [Muraki, 1991]
  - Who causes the event
  - Who is influenced by the event

Agentivity, Focus

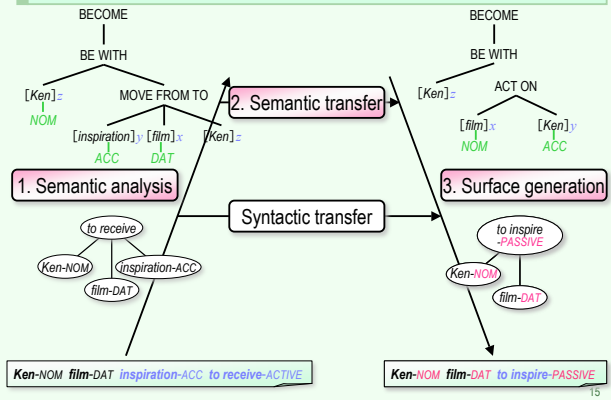
- Syntactic cases:
  - Which marker should be assigned for each nominal element

Linking

Contents

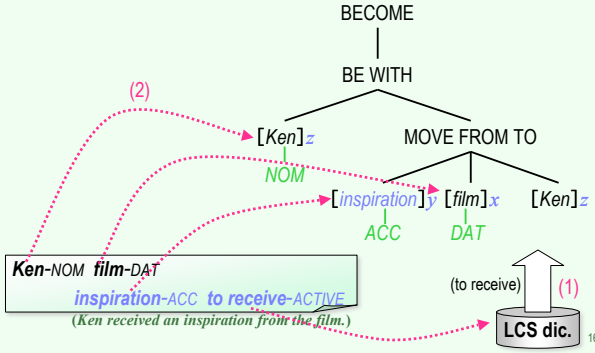
1. Issues and goals
2. LCS
3. Paraphrasing of LVCs in Japanese
4. LCS-based paraphrase generation model
5. Experiments
6. Conclusion

LCS as lexical constraints / a tool for transfer



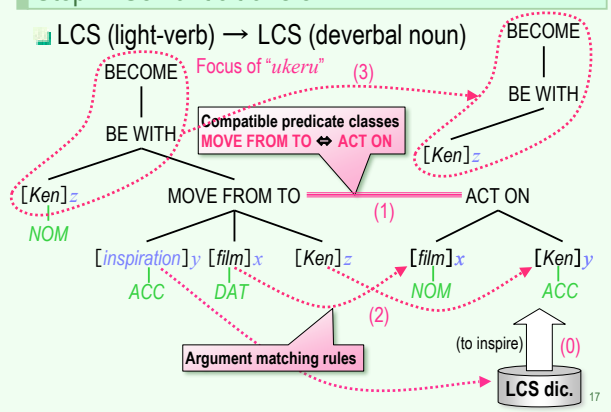
Step 1: Semantic analysis

- Surface → LCS (light-verb)



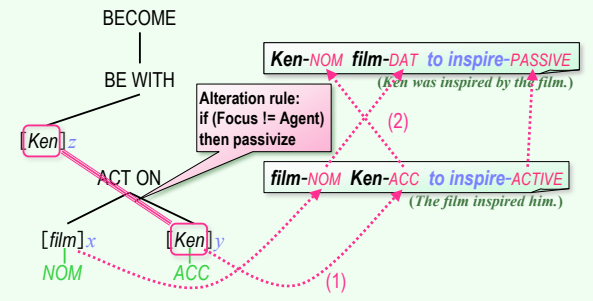
Step 2: Semantic transfer

- LCS (light-verb) → LCS (deverbal noun)



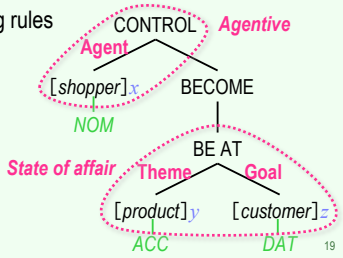
Step 3: Surface generation

- LCS (deverbal noun) → Surface



Summary of the rules for paraphrasing of LVCs

- Depth-first matching in LCS transformation
  - 2 predicate classes:
    - Agentive: **CONTROL**, **ACT ON**, **ACT**, etc.
    - State of affair: **MOVE TO**, **BE AT**, **BE WITH**, etc.
  - 2 argument matching rules
- Generation:
  - A decision list
    - Causativization * 2
    - Passivization * 1
    - Leave active * 2



Contents

1. Issues and goals
2. LCS
3. Paraphrasing of LVCs in Japanese
4. LCS-based paraphrase generation model
5. Experiments
6. Conclusion

Setting

- LCS dictionary
  - 1,165 deverbal nouns (T-LCS dic. ver. 0.95)
  - 40 frequent light-verbs (manually collected and assigned LCS)
- Gold-standard

  (1) 3 clauses for each of 245 most frequent types of LVC

  (2) annotators produced same paraphrases for 711 clauses in terms of determining Voice and Syntactic cases
- Models
  - LM (baseline): selects a combination of voice and syntactic cases
  - LCS (proposed): generates all semantically explainable candidates
  - LCS+LM: filters anomalies among the output of LCS

Results

	LM	LCS	LCS+LM
# of candidates	547	798	717
# of correct paraphrases	322	624	609
# of incorrect paraphrases	225	174	108
Recall	.453	.878	.857
Precision	.589	.782	.849
F-measure (α =0.5)	.512	.827	.852

- LM < LCS < LCS+LM
  - Lexical properties encoded in LCS are useful
  - LM itself does not work well, but contributes to filter out anomalies among semantically derived paraphrases

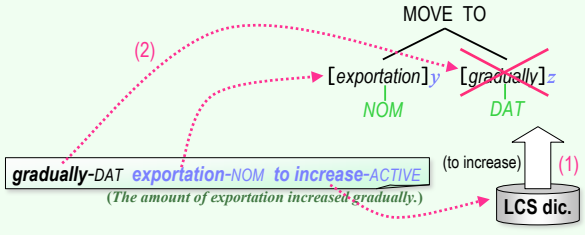
Error distribution

		LCS	LCS+LM
(1) Ambiguous role of dative	Step 1	78	47
(2) Transformation algorithm	Step 2	59	36
(3) Definition of LCS	Step 0	30	19
Other errors	-	7	6
Total	-	174	108

- Countermove
  - LCS typology should be firstly refined
  - For (1), semantic parsing is necessary
  - For (2), transformation principles should be reconsidered

Error source: ambiguous role of dative @ step 1

- Dative case often functions as an adjunct
  - Largest portion
  - Violation of selectional restriction for argument
  - We need a semantic parsing technology



Recent advances in semantic parsing

- Semantically annotated corpus / lexicon
  - FrameNet [Baker et al., 1998]
  - VerbNet [Kipper et al., 1998]
  - Propositional Bank [Palmer et al., 2005]
  - IAMTC [Dorr et al., 2004]
- Semantic parsing technology
  - Word sense disambiguation, semantic role labeling, etc.
  - CoNLL-2004, CoNLL-2005 Shared Task
  - Statistical methods have been well-discussed

Conclusion

- Exploiting LCS
  - Lexical constraints (syntactic and semantic properties)
    - Agentivity, Focus of statement, Linking
  - Tool for semantic transfer
- A model for paraphrasing of LVCs in Japanese
  - Small sets of linguistically explainable rules
  - F-measure: .512 (LM) < .814 (LCS) < .839 (LCS+LM)
  - Error analysis guides further research avenues

Current work

- Restructuring LCS dictionary [Takeuchi et al., 2005]
  - Re-organize the lexical properties to be encoded
  - Enlarge LCS dictionary
- Example-based semantic parsing [Hirano et al., 2005]
  - By collecting semantically labeled examples
  - Technical issue is to reduce human labor for labeling
- Enhancing LCS-based paraphrasing model
  - Predicate / argument matching algorithms
    - Induce by comparing source and target LCSs
  - Implementation for other classes of paraphrase

Future work

- Ultimate goal: cover various paraphrases
  - Harmonizing semantics-based paraphrasing with automatic paraphrase acquisition
    - 1. Build semantics-based paraphrase generation models for lexically compositional paraphrases
    - 2. Acquire paraphrases from corpus/Web
    - 3. Distill them into “idiosyncratic paraphrases” by decomposing them using models built in step 1