A hybrid approach toward Natural Language Understanding

Daisuke Bekki

(Joint work with Koji Mineshima,
Pascual Martinez-Gomez and Hitomi Yanaka)

Talk in Centre for Linguistic Theory and Studies in Probability (CLASP), December 09, 2020

Natural Language Understanding system “ccg2lambda”

- Topics
  - Recognizing Textual Entailment
  - Semantic Textual Similarity
  - FraCaS/JSeM test suits
- GitHub: https://github.com/mynlp/ccg2lambda

Computational Model of Natural Language Semantics based on Dependent Type Theory

Recognizing Textual Entailment (RTE)

T: Most new employees request a transfer to Osaka
H: Most new employees request something

entailment

T: Most new employees request a transfer
H: Most employees request a transfer

no entailment

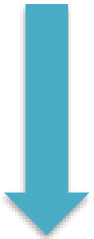
T: John failed to catch the 7 o'clock train
H: John caught the 7 o'clock train

contradiction

Syntactic/Semantic Parsing

- Mapping texts into semantic representations

Sentence Which city is the capital of Japan?

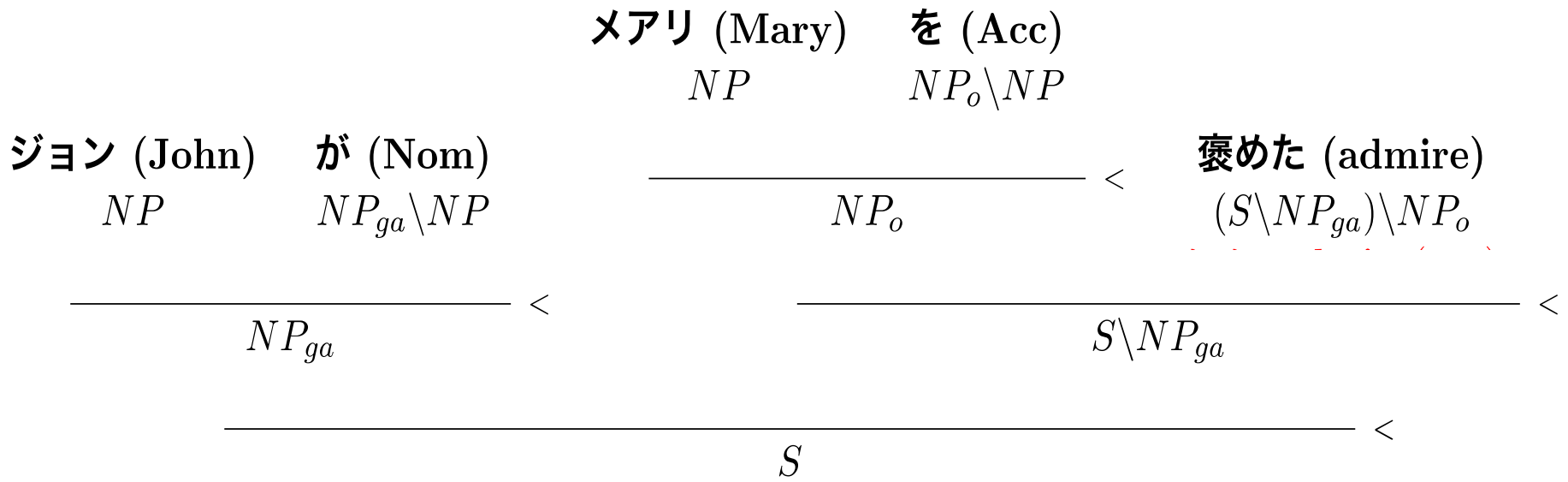


Grammar (CCG)

Meaning $\lambda x.\ city(x) \wedge capital(x, Japan)$

Combinatory Categorial Grammar (CCG)

- Open-domain CCG parsers
  - C&C [Clark and Curran, 2007]
  - EasyCCG [Lewis and Steedman, 2014]
  - depccg [Yoshikawa+, 2017]



Building CCG parsers

Noji and Miyao [ACL2016]
Yoshikawa+ [ACL2017]

English

Penn Treebank

CCGBank

CCG parser
- C&C parser
✓ **depccg**

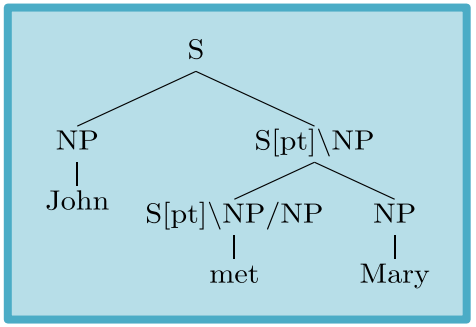
Japanese

Kyoto/NAIST Corpus

✓ **Japanese CCGBank**

✓ **CCG parser**
✓ **Jigg**
✓ **depccg**

```
(S (NP-SBJ-1 John)
   (VP (VBN met)
       (NP Mary)))
```



The first Japanese CCG parsers

Logical Inference

P : Smoking is prohibited in most cities.

$$\exists x(\text{smoking}(x) \wedge \text{most}(\lambda y.\text{city}(y), \lambda y.\text{prohibited}(x) \wedge \text{in}(x, y)))$$

Logic

H : Smoking is not allowed in some cities.

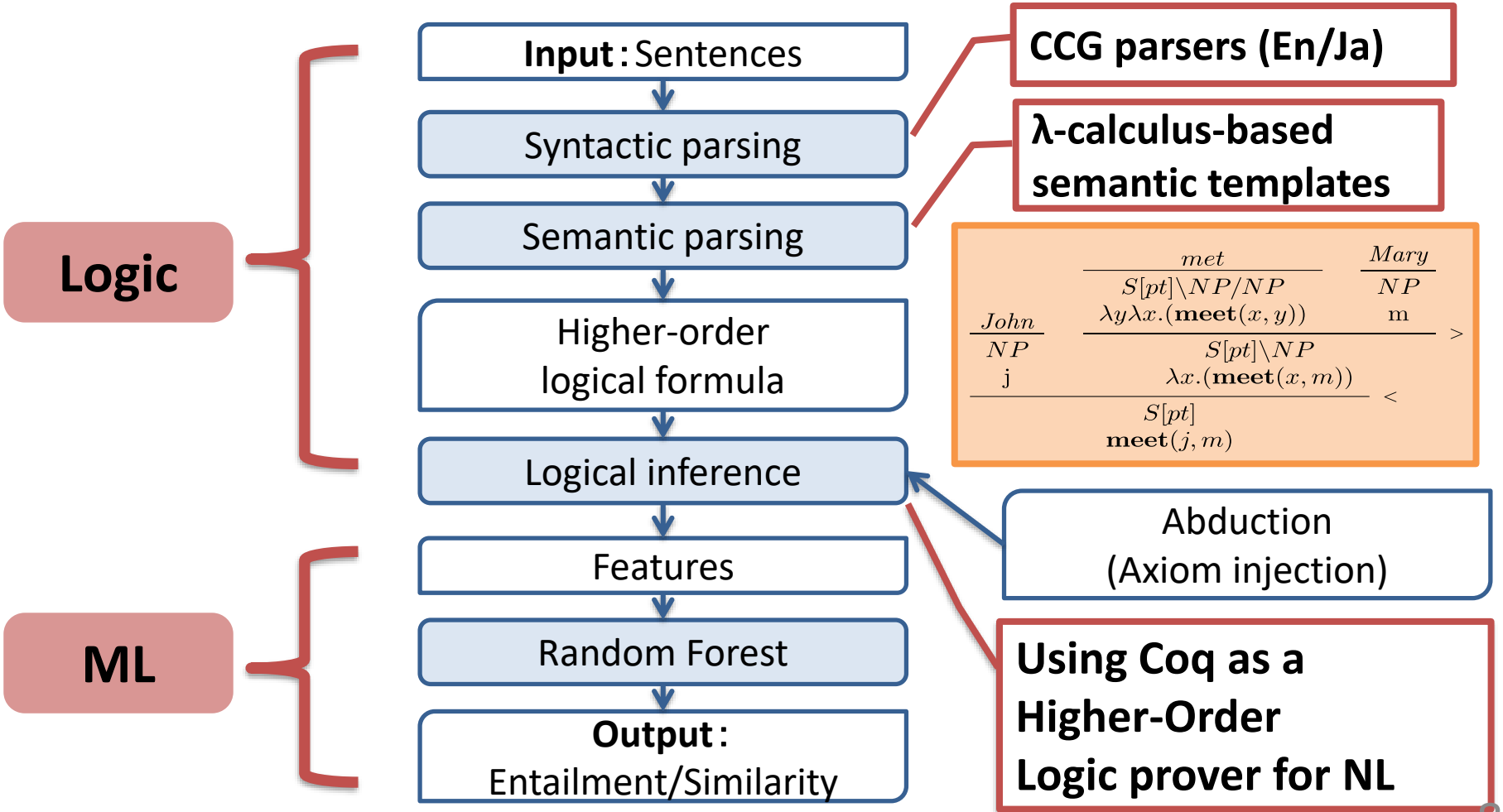
$$\exists x(\text{smoking}(x) \wedge \exists y(\text{city}(y) \wedge \neg\text{allowed}(x) \wedge \text{in}(x, y)))$$

ccg2lambda: Pipeline

https://github.com/mynlp/ccg2lambda

Mineshima+ [EMNLP2015, 2016]
Martínez-Goméz+ [ACL2016]
Yanaka+ [EMNLP2017, EMNLP2018]

End-to-end, open-domain semantic parser with inference system



Overview of FraCaS/JSeM project

FraCaS and MultiFraCaS

FraCaS test suite (Cooper et al. 1996):

`http://www-nlp.stanford.edu/~wcmac/downloads/fracas.xml`

An inference data set that

- covers core semantic phenomena
  - Generalized Quantifiers, Plurals, Nominal anaphora, Ellipsis, Adjectives, Comparatives, Temporal reference, Verbs, Attitudes
- requires minimal world knowledge
- is machine readable (McCartney and Manning 2007)
- has been used to evaluate NLP systems

MacCartney and Manning 2007, 2008	Lewis and Steedman 2013	Tian et al. 2014	Abzianidze 2014	Mineshima et al. 2015
Natural Logic	CCG & FOL	DCS	NL Tableau	CCG & HOL

MultiFraCaS: `http://www.ling.gu.se/~cooper/multifracas/`

- Translation of FraCaS test suite into Farsi, German, Greek, and Mandarin

FraCaS

1 GENERALIZED QUANTIFIERS

1.1 Conservativity

```
Q As are Bs == Q As are As who are Bs
```

fracas-001 answer: yes

P1	An Italian became the world's greatest tenor.
Q	Was there an Italian who became the world's greatest tenor?
H	There was an Italian who became the world's greatest tenor.

fracas-002 answer: yes

P1	Every Italian man wants to be a great tenor.
P2	Some Italian men are great tenors.
Q	Are there Italian men who want to be a great tenor?
H	There are Italian men who want to be a great tenor.
Note	*Note that second premise is unnecessary and irrelevant.*

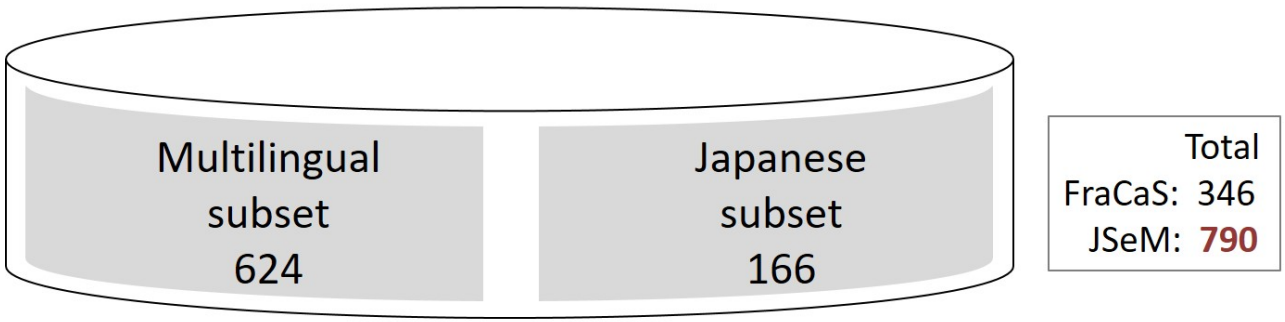
MultiFraCaS

fracas-001	lang: fa	answer: yes
P1		
script	یک ایتالیایی بزرگترین تنور جهان شد	
translit	yek italiayi bozorgtarin tenor e jahan shod.	
morph	An Italian greatest tenor of world became	
English	An Italian became the world's greatest tenor.	
Q		
script	آیا یک ایتالیایی وجود داشت که بزرگترین تنور جهان بشود؟	
translit	Aya yek italiayi vojood dasht ke bozorgtarin tenor e jahan beshavad?	
morph	Question-word one Italian there was who[that] greatest tenor of world become[subj 3sg] ?	
English	Was there an Italian who became the world's greatest tenor?	
H		
script	یک ایتالیایی وجود داشت که بزرگترین تنور جهان شد	
translit	yek italiayi vojood-dasht ke bozorgtarin tenor e jahan shod.	
morph	an Italian there-was[3sg] who[that] greatest tenor of world became[3sg]	
English	There was an Italian who became the world's greatest tenor.	
A		
script	آری	
translit	Ari	
English	Yes	
Note		

JSeM test suite

JSeM (Kawazoe et al. 2015)

`http://researchmap.jp/community-inf/JSeM/`



Multilingual subset
- Japanese counterparts of FraCaS problems (cf. MultiFraCaS project)

Japanese subset
- Universal phenomena not covered by FraCaS e.g. modality, conditionals, adverbs, focus
- Japanese-specific phenomena

JSeM test suite

jsem-id:1	answer: yes	inference type: entailment	phenomena: generalized quantifier, conservativity
	linked to: fracas-001	literal translation?: yes	same phenomena?: unknown

P1
script ある イタリア 人 が 世界 最高 の テノール 歌手 に なった 。
English An Italian became the world's greatest tenor.
H
script 世界 最高 の テノール 歌手 に なった イタリア 人 が いた 。
English There was an Italian who became the world's greatest tenor.

jsem-id:2	answer: yes	inference type: entailment	phenomena: generalized quantifier, conservativity, Q-no NC
	linked to: fracas-001	literal translation?: yes	same phenomena?: unknown

P1
script 一人 の イタリア 人 が 世界 最高 の テノール 歌手 に なった 。
English An Italian became the world's greatest tenor.
H
script 世界 最高 の テノール 歌手 に なった イタリア 人 が いた 。
English There was an Italian who became the world's greatest tenor.
Note

“Inference as Tests” paradigm

Proposal

Why don't we evaluate frameworks/analyses/hypotheses of formal/computational semantics (MG, H&K, DRT, DPL, MRS, CS, TTR, DTS, etc) by using FraCaS/JSeM?

Motivations

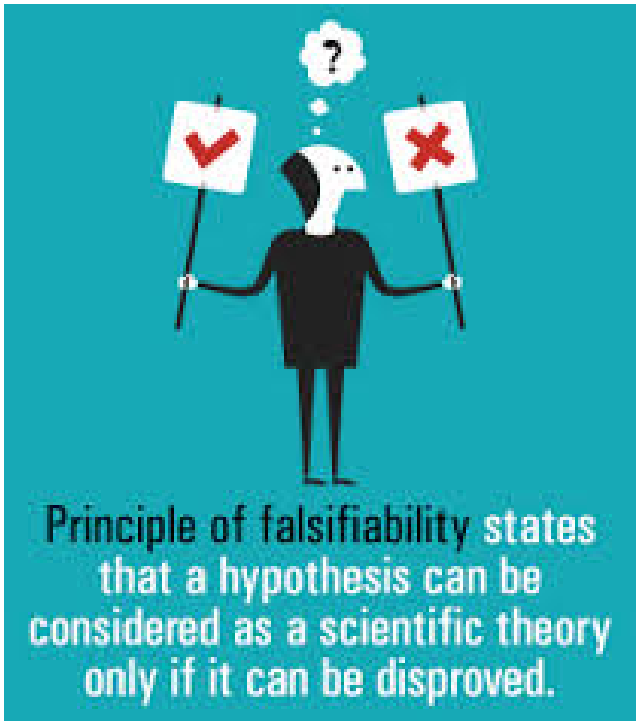
1. Ensuring falsifiability of semantic theories
2. Evaluating A.I. systems
3. Preserving our semantic knowledge

Imagine

- Every semantic paper accompanies linguistic data it covers in FraCaS/JSeM format
- Data are checked their reproducibility, and eventually be added to FraCaS/JSeM test suites
- Some frameworks are ensured to provide a good analysis for most data in FraCaS/JSeM

https://www.flickr.com/photos/vincgalery/7471905776

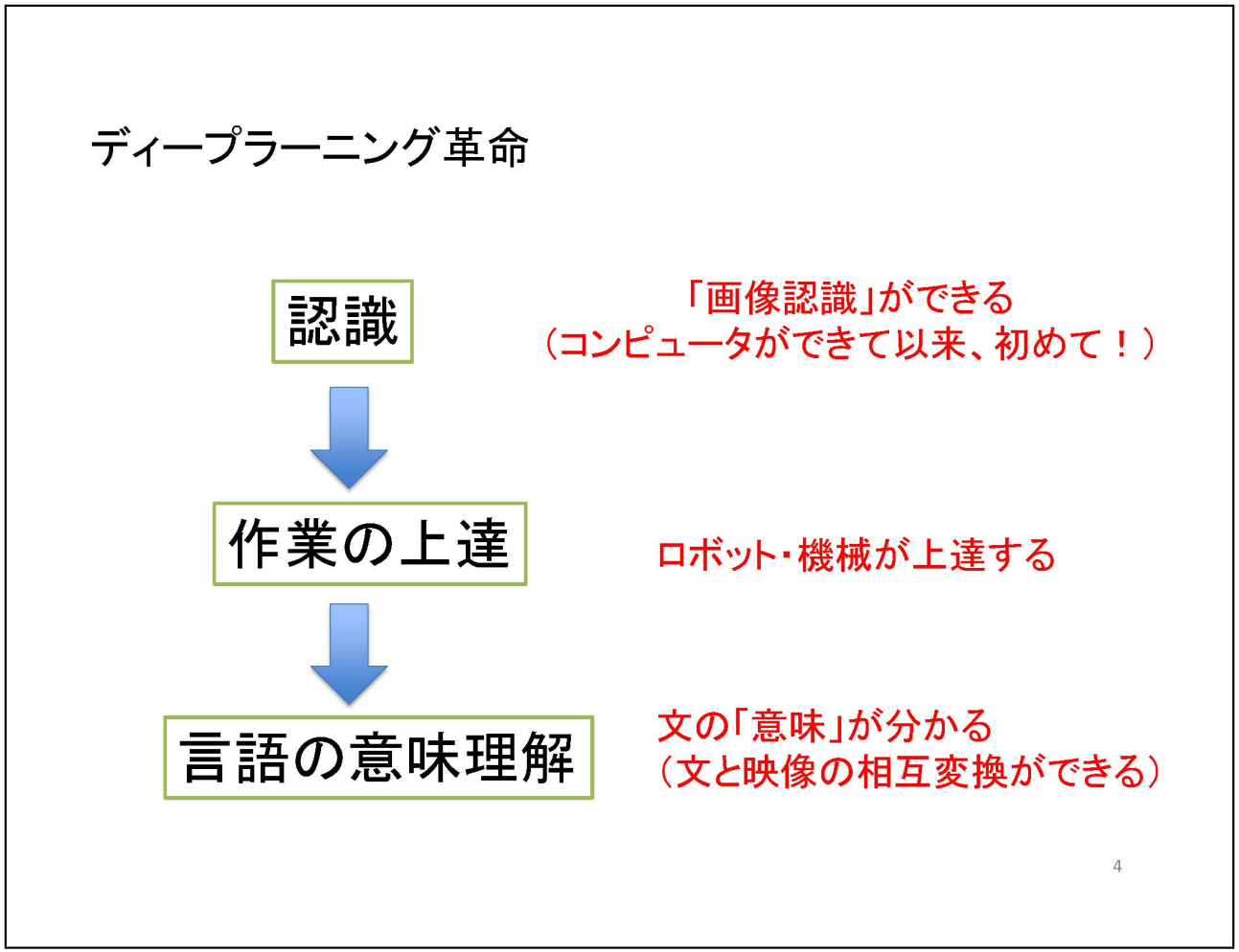
1. Ensuring falsifiability of semantic theories



http://www.buzzle.com/articles/principle-of-falsifiability.html

Falsifiability of most semantic theories are unclear due to the absence of the way to prove that every theoretical notion is well-defined so that they actually yield empirical predictions.
Implimentation is a way to make clear the falsifiability of a given theory.

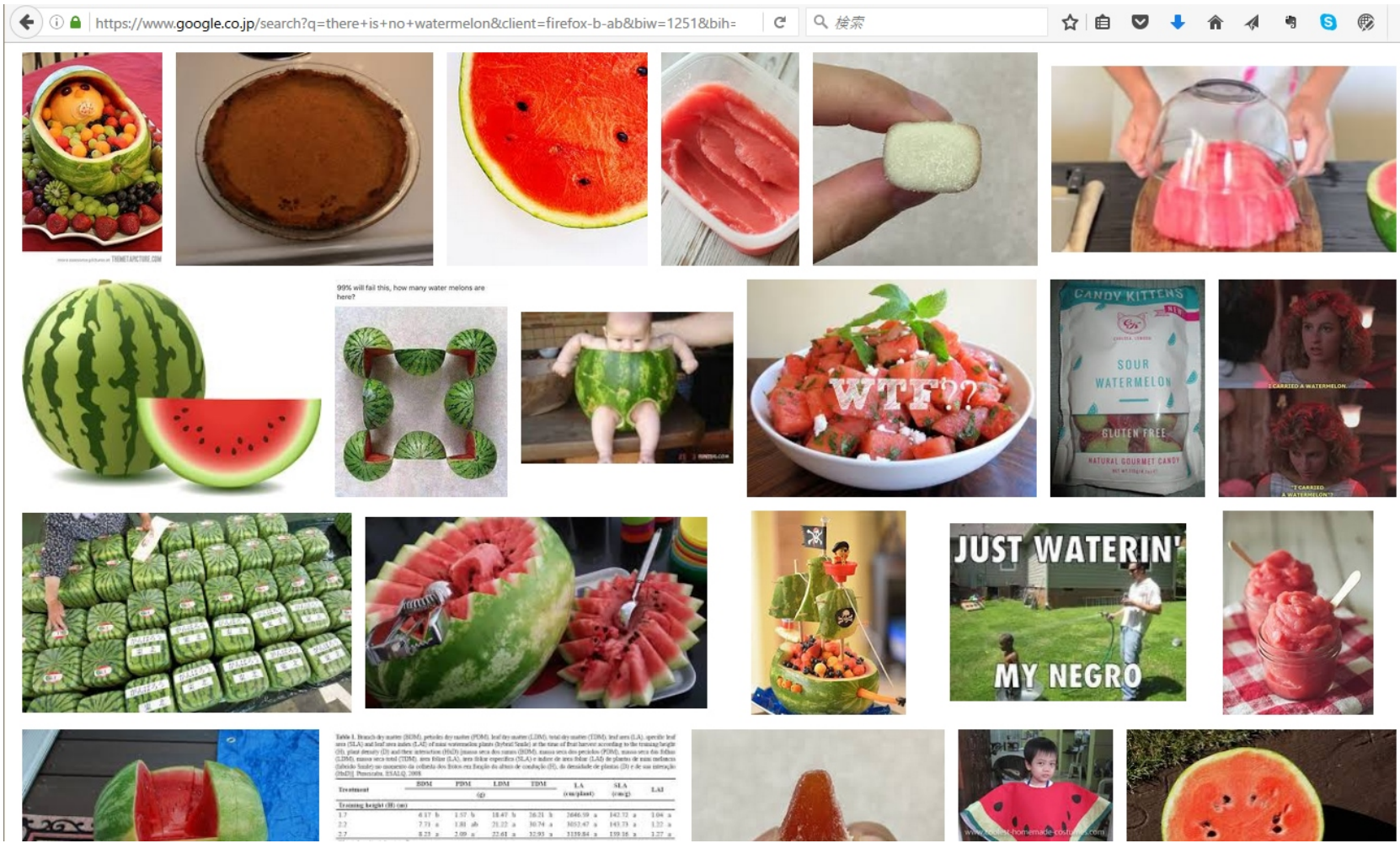
2. Evaluating A.I. systems



http://www.mof.go.jp/pri/research/conference/fy2016/inv_01_02.pdf

"Understanding the meaning of a sentence
= Translation of a sentence into an image" (!)

“There is no watermelon.”



2. Evaluating A.I. systems

"Understanding the meaning of a sentence" involves understanding of:

- Generalized Quantifiers
- Plurals
- Nominal anaphora
- Ellipsis
- Adjectives
- Comparatives
- Temporal reference
- Verbs
- Attitudes
- ...

http://www.senken.co.jp/news/corporation/sato-pepper-160722/

Existing inference data sets

For English

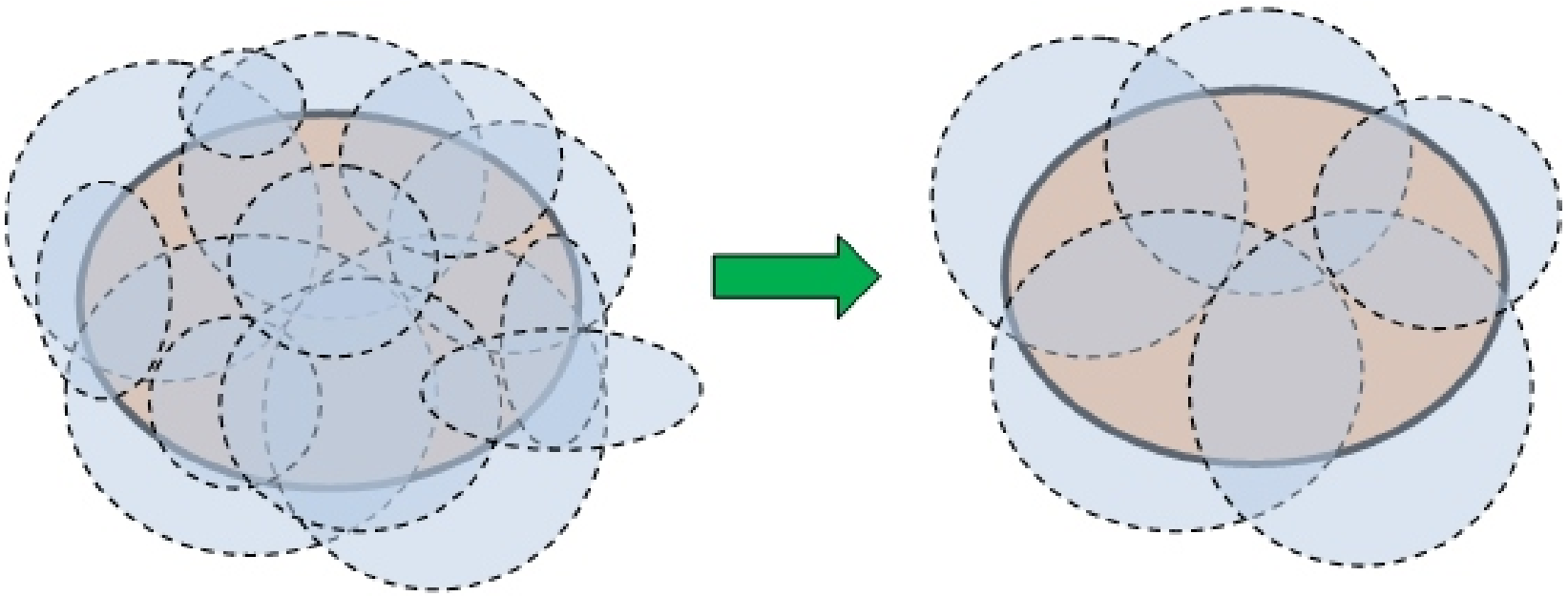
- PASCAL RTE data sets (Dagan et al., 2006)
- SemEval data sets (Marelli et al., 2014)(SICK - Sentences Involving Compositional Knowledge)
- Stanford Natural Language Inference corpus (Bowman et al., 2015)

NLI

For Japanese

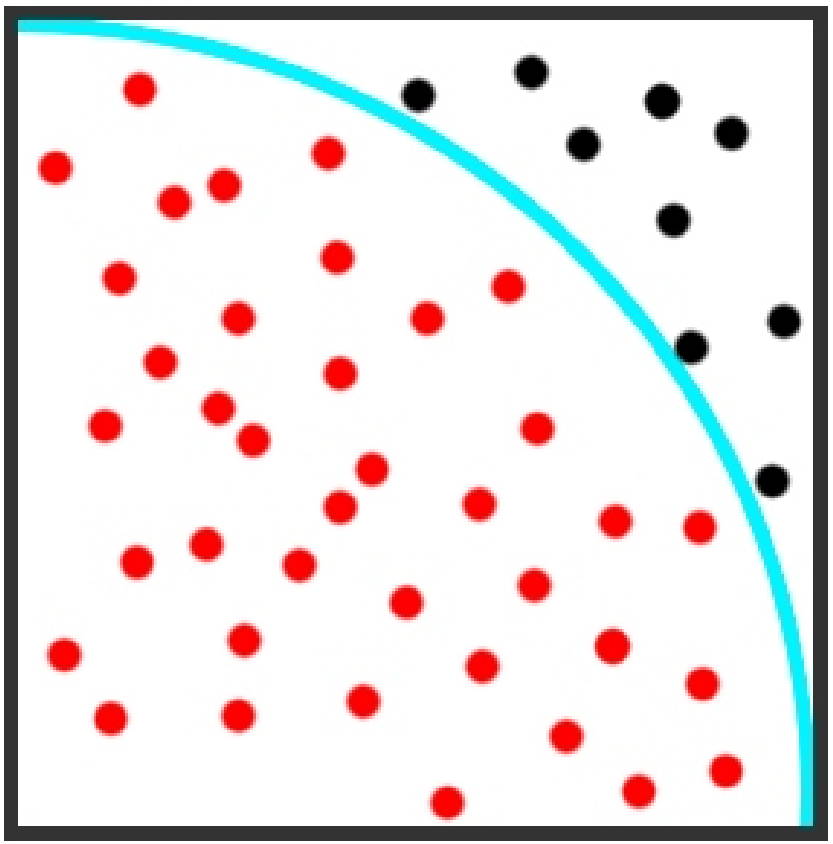
- NTCIR RITE data sets (2012-)
- Kyoto Univ. Textual Entailment test data (Kotani et al., 2008)

3. Preserving our semantic knowledge



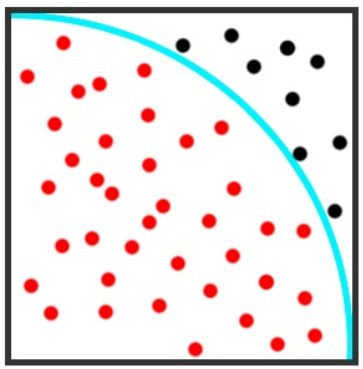
https://simomaths.wordpress.com/2013/02/24/topology-sequentially-compact-spaces-and-compact-spaces/

How theoretical linguists provide linguistic data



http://nigohiroki.hatenablog.com/entry/2012/09/18/232306

Misunderstanding about linguists



- "Linguists only deals with the data they are interested in."
- "Linguists only deals with sentences that we never say in real life."

"Inferences as Tests" paradigm

"Inferences as Tests" paradigm

Semantic frameworks (theories and implementations) are required to correctly predict validities of inferences that have been shown to be reproducible.

An inference pattern of a sentence is not its meaning by itself; but it serves as a test set to verify/falsify semantic theories/analyses/hypotheses about its meaning.

Discussions

Two questions about "Inferences as Tests" paradigm

1. Is entailment only a small part of semantic phenomena?
2. Can all the semantic phenomena be described as textual entailments?

Is entailment only a small part of semantic phenomena?

No.

Semantic Anomaly (on the basis of contradiction)

Heim and Kratzer (1998): "Quantifying expressions are not of type e"

(1) # Taro is Japanese, and Taro is not Japanese.

(2) Someone is Japanese, and someone is not Japanese.

H Taro is Japanese, and Taro is not Japanese.
answer: NO

H Someone is Japanese, and someone is not Japanese.
answer: UNKNOWN

Presupposition

That *Mary takes care of John's dog* presupposes *John has a dog* can be expressed in the form of family of sentences tests (Kadmon, 2001)

P	Mary takes care of John's dog.
H	John has a dog.
answer:	YES

P	Mary does not take care of John's dog.
H	John has a dog.
answer:	YES

P	If Mary takes care of John's dog, John is happy.
H	John has a dog.
answer:	YES

Presupposition

Filter

P	If John has a dog, Mary takes care of John's dog.
H	John has a dog.
answer:	UNKNOWN

Plug

P	Susan saids that Mary takes care of John's dog.
H	John has a dog.
answer:	UNKNOWN

Can all the semantic phenomena be described as textual entailments?

No.

Japanese Honorification

P	Sam-ga O-warai-ninat-ta.
	Sam-NOM subj.hon-laugh-subj.hon-PAST
	'Sam laughed.'
H	The speaker honors Sam.
answer:	YES

- Is it true that the speaker honors Sam?
- Expressives (honorifics, discourse particles, etc.): can we articulate their meaning? although their usages are relatively clear (Kaplan 1999; Potts 2003, a.o)

Conversational Implicatures

P	Mary was born in Osaka or Kyoto.
H	The speaker does not know whether Mary was born in Osaka or Kyoto.
answer:	YES??

- Cancellable implicatures should be distinguished from entailments.
- How can the data about conversational implicatures integrated into FraCaS/JSeM?

Accomodation

Familiality condition:

(3) John found a rabbit, but the animal run away.

(4) # John found an animal, but the rabbit run away.

- The status of the latter sentence is controvertial:
  - (4) is out / can be accepted via accommodation / totally ok.
  - If accommodation is a part of our semantic competence, how can we describe the difference between (3) and (4)?

Law of Excluded Middle

Heim and Kratzer (1998):

(5) Taro is taller than John, or Taro is equal or smaller than John.

(6) Someone is taller than John, or someone is equal or smaller than John.

H Taro is taller than John,
or Taro is equal or smaller than John.
answer: YES??

H Someone is taller than John,
or someone is equal or smaller than John.
answer: UNKNOWN

Resoucefulness

Right node raising:

(7) John loves and Bill hates, Susan.

- In Chomsky (1957), (7) was judged as unacceptable, but its status has been changed into "acceptable" in the past 60 years.
- If we ever change the status of a certain item in FraCaS/JSeM, should we change the evaluations of systems based on it retrospectively?

Future Perspectives

A Research Program of Formal Semantics based on "Inferences as Tests" paradigm

"Formal syntax/semantics framework" competition to evaluate the performance of each system in terms of:

- the number of problems solved
- the runtime for problems solved

This might be far better than "In our analysis, we adopt Heim and Kratzer since it is the standard framework..." How many FraCaS problems can the framework that you adopt in your paper solve?

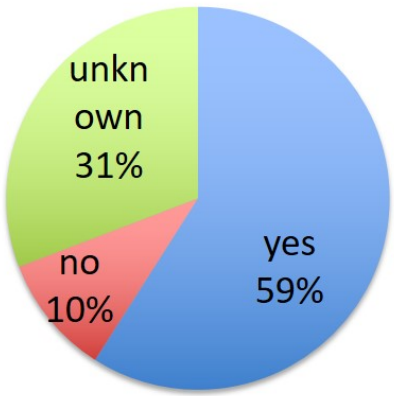
A Research Program of Formal Semantics based on "Inferences as Tests" paradigm

A number of wide-coverage "formal semantics" systems have been implemented recently:

- Bos et al. (2004, Boxer, CCG + DRT) for English
- Moot (2010, Grail, TLG + DRT) for French
- Butler and Yoshimoto (2012, SCT + Treebank Semantics) for English and Japanese
- Tian et al. (2014, DCS) for English GQ
- Abzianidze (2015, EMNLP, CCG + NL Tableau) for English (Sick Dataset)
- Mineshima et al. (2015, EMNLP, ccg2lambda = CCG + HOL/DTS) for English and Japanese (FraCaS, JSeM and SICK) (`https://github.com/mynlp/ccg2lambda`)
- Chatzikyriakidis et al. (2015, Coq)

Remaining Problems

- The data is not well-balanced (yes-no-unknown)



- Some sections have very few examples

	GQ	plural	anaphora	ellipsis	adjectives	comparatives	temporal	verbs	attitudes
count	80	33	28	55	23	31	75	8	13
%	23	10	8	16	7	9	22	2	4

- Missing important phenomena (modality, conditionals, negation and many others)

JSeM on sale

- JSeM project: the Japanese inference data set, a benchmark for formal/computational semantics and NLP systems, part of which serves as a Japanese MultiFraCaS
- β-version is released:
  http://researchmap.jp/community-inf/JSeM/
- Necessity:
  1. Ensuring falsifiability of semantic theories
  2. Evaluating A.I. systems
  3. Preserving our semantic knowledge

Reference I

Abzianidze, L. (2015) "Towards a Wide-coverage Tableau Method for Natural Logic", In: T. Murata, K. Mineshima, and D. Bekki (eds.): *New Frontiers in Artificial Intelligence: JSAIisAI 2014 Workshops, LENLS, JURISIN, and GABA, Revised Selected Papers. Lecture Notes in Computer Science, volume 9067*. pp.66–82.

Bos, J., S. Clark, M. J. Steedman, J. R. Curran, and J. Hockenmaier. (2004) "Wide-Coverage Semantic Representations from a CCG Parser", In the Proceedings of *COLING '04*. Geneva.

Bowman, S., G. Angeli, C. Potts, and C. D. Manning. (2015) "A large annotated corpus for learning natural language inference".

Butler, A. and K. Yoshimoto. (2012) "Banking meaning representations from treebanks", *Linguistic Issues in Language Technology (LiLT)* **7**(6), pp.1–22.

Reference II

Chomsky, N. (1957) *Syntactic Structure*. Mouton the Hague.

Cooper, R., D. Crouch, J. van Eijck, C. Fox, J. van Genabith, J. Jaspars, and K. Konrad. (1996) "Using the framework", Technical report.

Dagan, I., O. Glickman, and B. Magnini. (2006) "The PASCAL recognising textual entailment challenge", In: *Machine learning challenges. evaluating predictive uncertainty, visual object classification, and recognising tectual entailment*. Springer Berlin Heidelberg, pp.177–190.

Heim, I. and A. Kratzer. (1998) *Semantics in Generative Grammar*. Malden, MA., Blackwell Publishers.

Kadmon, N. (2001) *Formal Pragmatics*. Blackwell.

Reference III

Kotani, M., T. Shibata, T. Nakata, and S. Kurohashi. (2008) "Building Textual Entailment Japanese Data Sets and Recognizing Reasoning Relations Based on Synonymy Acquired Automatically", In the Proceedings of *the 14th Annual Meeting of the Association for Natural Language Processing*. Tokyo, Japan.

Marelli, M., S. Menini, M. Baroni, L. Bentivogli, R. Bernardi, and R. Zamparelli. (2014) "A SICK cure for the evaluation of compositional distributional semantic models", In the Proceedings of *LREC*. pp.216–223.

Reference IV

Mineshima, K., P. Martinez Gomez, Y. Miyao, and D. Bekki. (2015) "Higher-order logical inference with compositional semantics", In the Proceedings of *Conference on Empirical Methods in Natural Language Processing (EMNLP2015)*. Lisboa, Portugal, pp.2055–2061.

Moot, R. (2010) "Wide-coverage French syntax and semantics using Grail", In the Proceedings of *TALN 2010*.

Sutcliffe, G. (2009) "The TPTP problem library and associated infrastructure. The FOF and CNF Parts, v3.5.0.", *Journal of Automated Reasoning* **43**(337).