How to Read/Write an International Conference Paper

Graham Neubig
Nara Institute of Science and Technology (NAIST)
2015-3-16

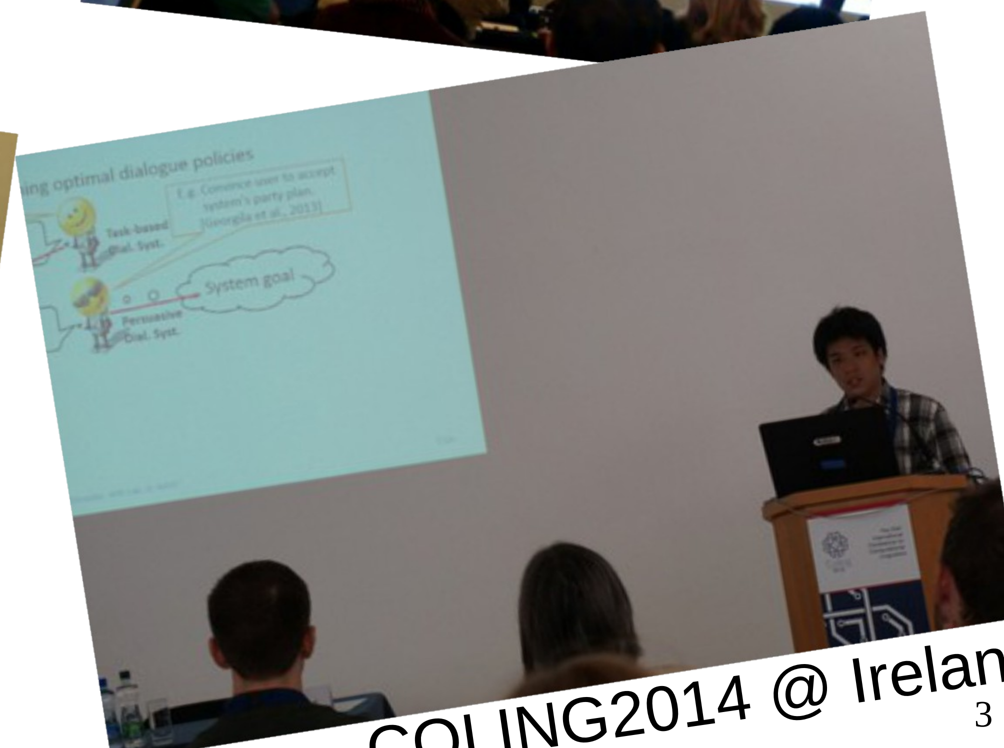
Tutorial at the Japanese Association for NLP 2015

Why do we Write?

COLING2014 @ Ireland

IWSDS2014 @ Korea

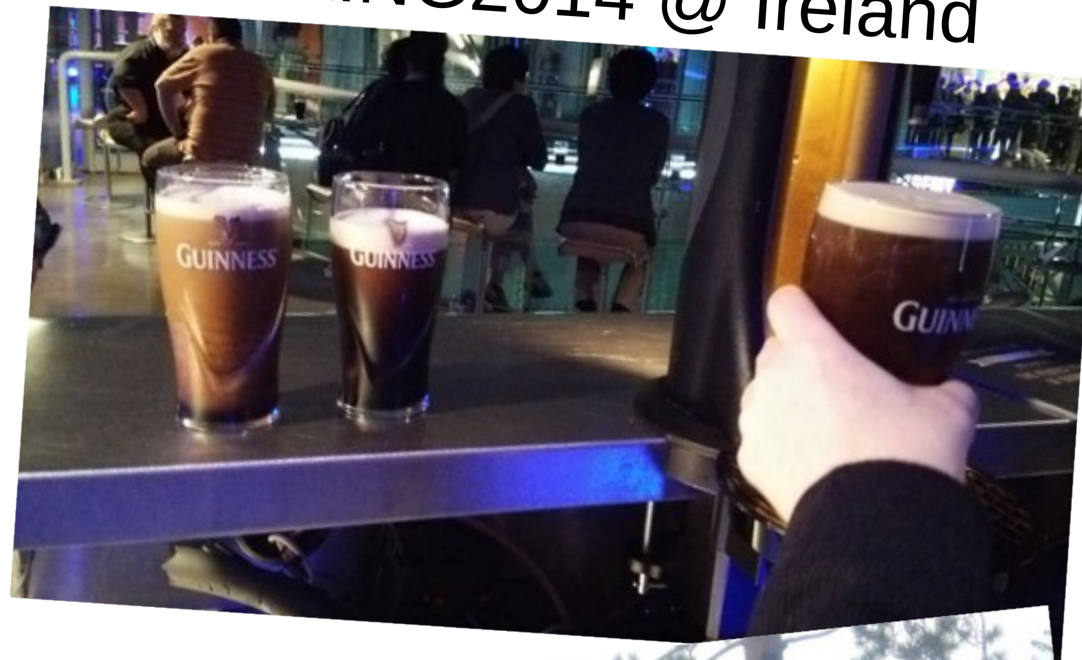
ACL2014 @ USA



COLING2014 @ Ireland

EACL2014 @ Sweden

COLING2014 @ Ireland



EMNLP2014 @ Qatar

IWSLT/SLT2014 @ USA

ACL Anthology
A Digital Archive of Research Papers in Computational Linguistics

Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)

P14-2089 [bib] [MRF: latexml]: **Mo Yu; Mark Dredze**
Improving Lexical Embeddings with Semantic Knowledge

P14-2090 [bib]: **Yusuke Oda; Graham Neubig; Sakriani Sakti; Tomoki Toda; Satoshi Nakamura**
Optimizing Segmentation Strategies for Simultaneous Speech Translation

P14-2091 [bib]: **Taku Kudo; Hiroshi Ichikawa; Hideto Kazawa**
A joint inference of deep case analysis and zero subject generation for Japanese-to-English statistical machine translation

SLT2014 @ USA

APSIPA2014 @ Cambodia

Where do we Submit?

If you want many people to read your paper:

1) Top international conferences
2) Workshops affiliated with top conferences
3) Others

For your first paper,
don't worry too much.

	Google Scholar	CORE Score	Rate
ACL Association for Computational Linguistics	62	A+	26%
EMNLP Empirical Methods in Natural Language Processing	53	A	27%
NAACL North American Chapter of the Association for Computational Linguistics	48	A	30%
COLING International Conference on Computational Linguistics	31	A	31%
EACL European Chapter of the Association for Computational Linguistics	30	A	22%
IJCNLP International Joint Conference on Natural Language Processing	15	B	36%

This is what you're up against...

However...

- There's no way I could do it...

  You'll never know if you don't try. Go for it!
- No-one will appreciate this work...

  That's for the reviewers to decide. Go for it!
- My work is not done. I want to finish it first...

  There is no “finished” research. Go for it!

**But when you do go for it,
do it with the best paper possible!**

Outline

- What is a “good” paper?
- The paper writing process
- Survey
- Paper structure, and each section
- Proofreading
- Basic English for research papers
- After acceptance

What is a “good” paper?

Definitions of Good Papers

- A paper that influences many people
- A paper that reviewers like

These are not equal!

“When you try to do something new, your paper will often get rejected. In fact, many of my papers that have won prizes have been rejected at some point.”

--An Anonymous Professor

Review Categories

- Clarity: Is it easy to understand?
- Novelty: Is it new?
- Meaningful Comparison: Does it compare well with previous work?
- Reliability: Are equations and experiments correct?
- Impact: Will it make a big difference in the field?
- Replicability: Could others replicate the experiments?
- Overall Evaluation: What did you think?

In the end, this is what matters.

What Decides Overall Evaluation?

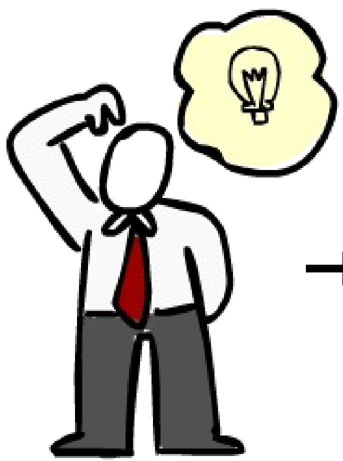
Could you tell your story?
(The problem, the solution)

Was it convincing?

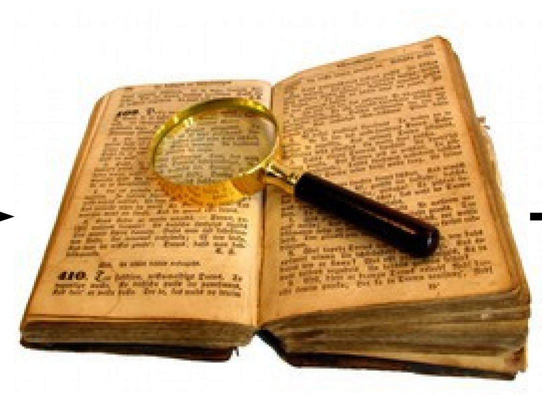
The Paper Writing Process

The Standard Process

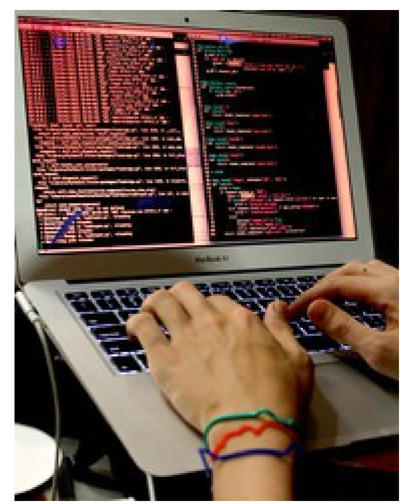
Idea



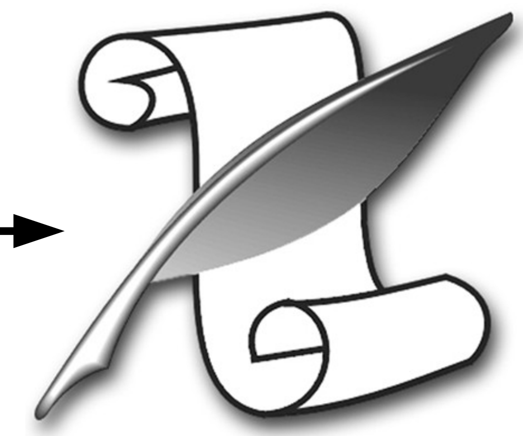
Survey



Experiments

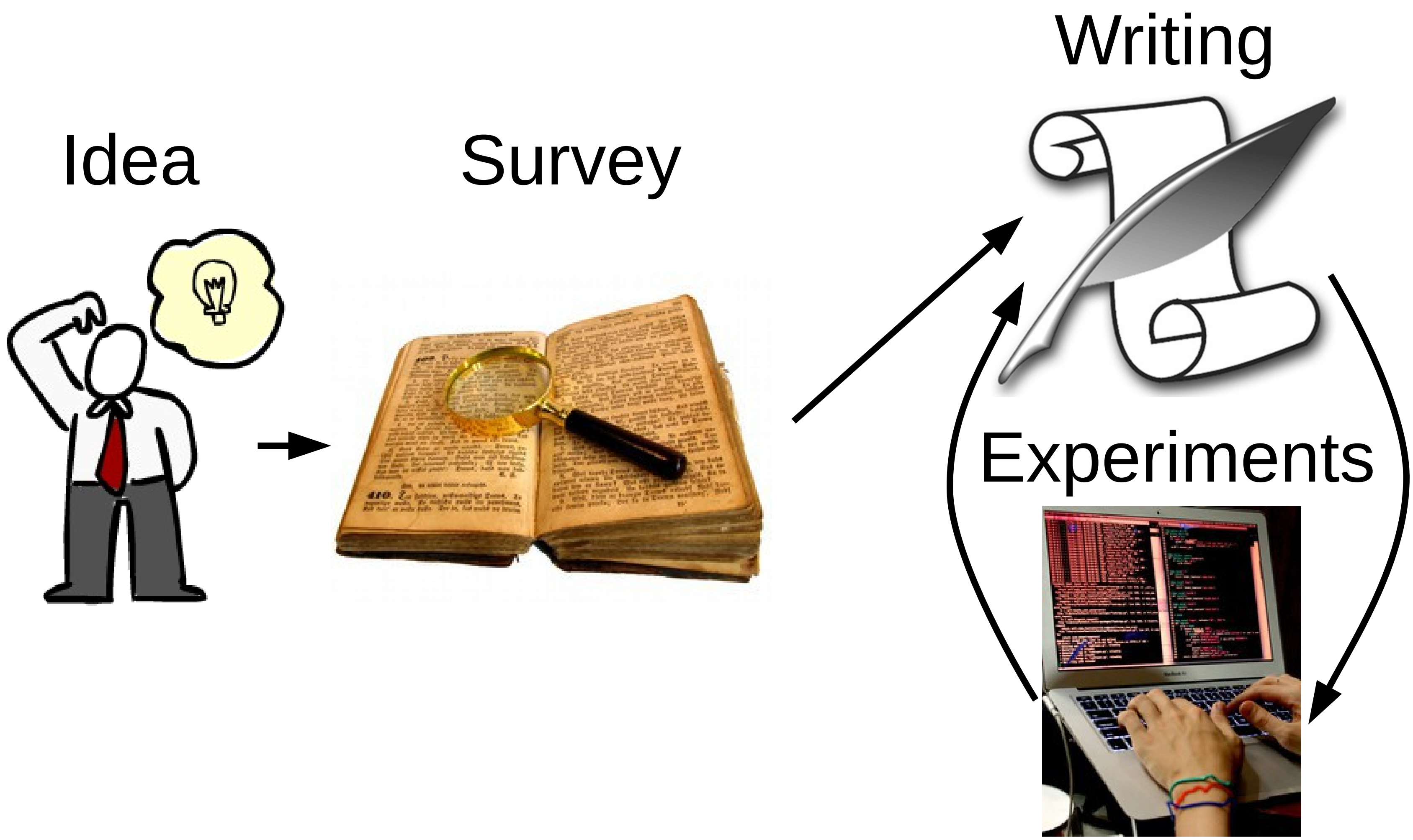


Writing



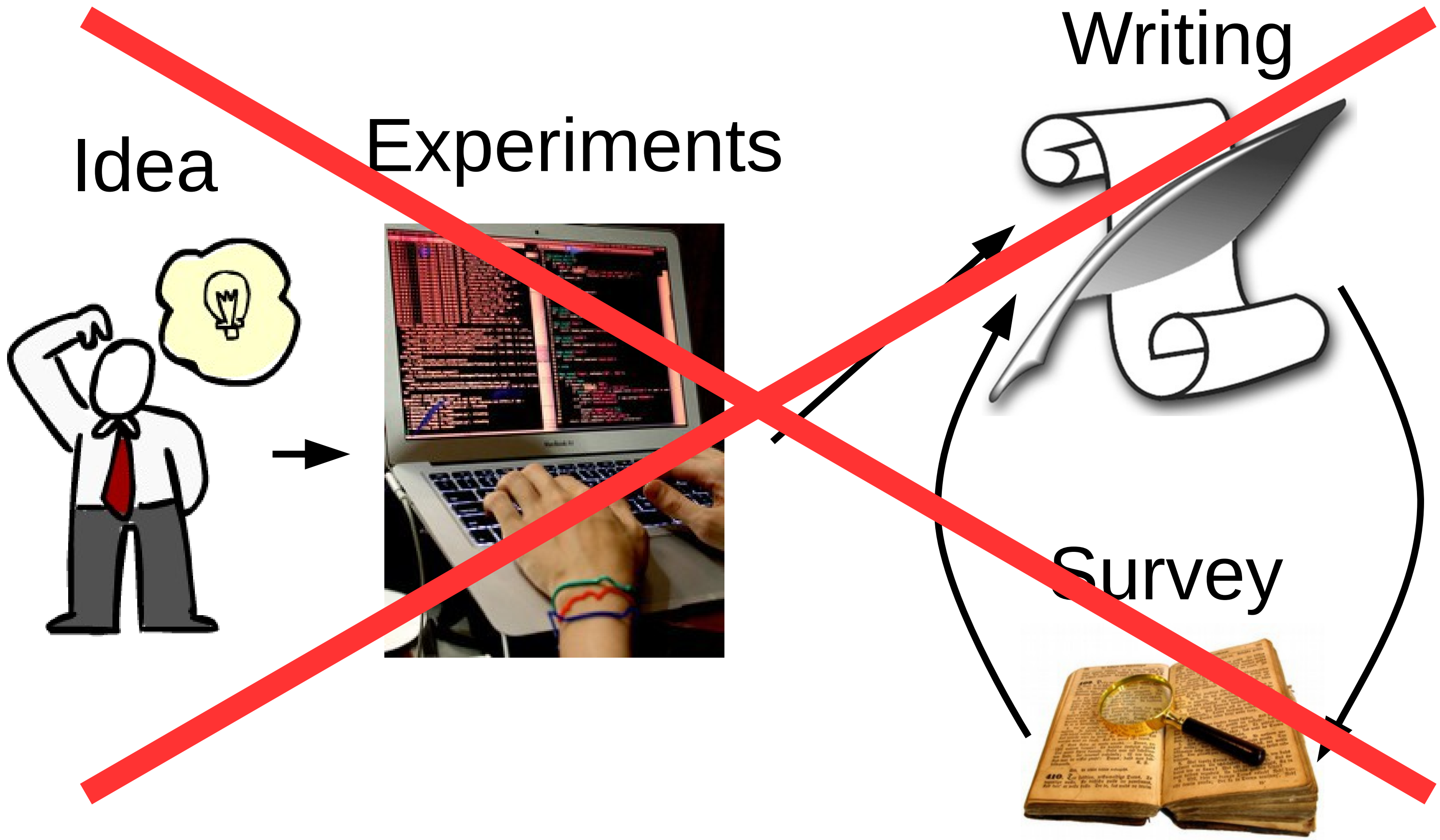
Images: Wikimedia Pictofigo, flickr HackNY.org, flickr Reinis Ivanovs

The "Write the Paper First" Process



Images: Wikimedia Pictofigo, flickr HackNY.org, flickr Reinis Ivanovs

The “I Just Can't Wait!” Process



Images: Wikimedia Pictofigo, flickr HackNY.org, flickr Reinis Ivanovs

Survey

How Many Papers should I Read?

Quiz:

How many papers are must be read for a good survey?

a) 10	b) 30	c) 100	d) 300	e) 1000
Not Enough	Good	Better!	Better!	Better!

The Survey Process

- Keyword search
- Find older/newer papers
- Read the abstract/intro
- Read details of the most related papers

Sources of Papers in Natural Language Processing

ACL Anthology

ACL Anthology
A Digital Archive of Research Papers in Computational Linguistics

http://www.aclweb.org/anthology/

Google Scholar

Google scholar

http://scholar.google.com/

ACL Anthology

- Covers most prestigious conferences/journals in NLP

ACL events

CL: Intro FS MT&CL 74-79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 00 01 02 03 04 05 06 07 08 09 10 11 12 13 UPDATED 14

TACL: NEW 15 14 13

ACL: Intro 79 80 81 82 83 84* 85 86 87 88 89 90 91 92 93 94 95 96 97* 98* 99 00 01 02 03 04 05 06* 07 08* 09* 10 11 12 13 14

EACL: Intro 83 85 87 89 91 93 95 97* 99 03 06 09 12 14

NAACL: Intro 00* 01 03 04 06* 07* 09* 10* 12* 13*

EMNLP: 96 97 98 99 00 01 02 03 04 05 06 07* 08 09 10 11 12* 13 14

CoNLL: 97 98 99 00 01 02 03 04 05 06 07 08 09 10 11 12 13 14

***Sem/ SemEval:** 98 01 04 07 10 12 13 14

ANLP: Intro 83 88 92 94 97 00*

Workshops: 90 91 93 94 95 96 97 98 99 00 01 02 03 04 05 06 07 08 09 10 11 12 13 UPDATED 14

SIGs: ANN BIOMED DAT DIAL FSM GEN HAN HUM LEX MEDIA MOL MT NLL PARSE MORPHON SEM SEMITIC SLPAT WAC

Other Events

COLING: 65 67 69 73 80 82 84* 86 88 90 92 94 96 98* 00 02 04 06* 08 10 12 14

HLT: 86 89 90 91 92 93 94 01 03* 04* 05 06* 07* 08* 09* 10* 12* 13*

IJCNLP: 05 08 09* 11 13

LREC: 00 02 04 06 08 10 12 14

PACLIC 95 96 98 99 00 01 02 03 04 05 06 07 08 09 10 11 12 13

Rocling Intro 88 89 90 91 92 93 94 95 96 97 98 99 00 01 02 03 04 05 06 07 08 09 10 11 12 13 NEW 14

TINLAP: 75 78 87

Donors Needed: COLING-65, any missing COLING

ALTA Intro 0

RANLP 09 11 1

JEP/TALN/RECITAL 12 13 1

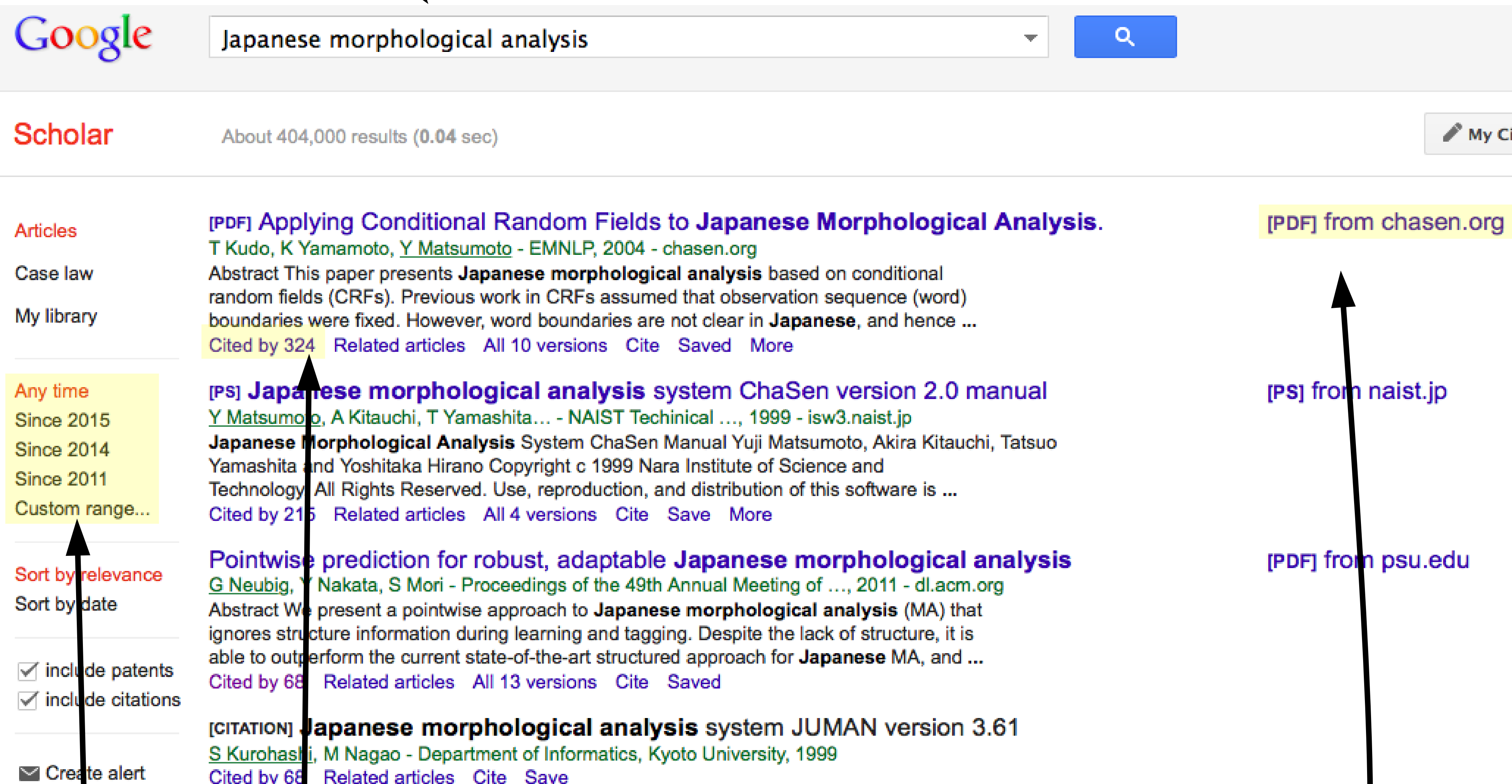
MUC: 91 92 9

Tipster: 93 96 9

In Progress: Finite S

- Start with past 5 years of ACL, NAACL, EMNLP, TACL

Search → Google Scholar



Years # of Citations Get PDFs

Finding Newer Papers

- Click “Cited By ...” in Google Scholar

[PDF] Applying Conditional Random Fields to **Japanese Morphological Analysis**.
T Kudo, K Yamamoto, Y Matsumoto - EMNLP, 2004 - chasen.org
Abstract This paper presents **Japanese morphological analysis** based on conditional random fields (CRFs). Previous work in CRFs assumed that observation sequence (word) boundaries were fixed. However, word boundaries are not clear in **Japanese**, and hence ...
Cited by 324 Related articles All 10 versions Cite Saved More

Gives a list of citing papers

An introduction to conditional random fields for relational learning
C Sutton, A McCallum - Introduction to statistical relational …, 2006 - books.google.com
Conditional random fields (CRFs) combine the modeling flexibility of graphical models with the ability to use rich, nonindependent features of the input. In this tutorial, we review modeling, inference, and parameter estimation in CRFs, both on linear chains and on ...
Cited by 716 Related articles All 3 versions Cite Save

[PDF] An introduction to conditional random fields
C Sutton, A McCallum - Machine Learning, 2011 - research.ed.ac.uk
Abstract Many tasks involve predicting a large number of variables that depend on each other as well as on other observed variables. Structured prediction methods are essentially a combination of classification and graphical modeling. They combine the ability of ...
Cited by 145 Related articles All 71 versions Cite Save More

[BOOK] Computational approaches to morphology and syntax
B Roark, RW Sproat - 2007 - anthology.aclweb.org
Brian Roark and Richard Sproat have written a compact and very readable book surveying computational morphology and computational syntax. This text is not introductory; instead, it will help bring computational linguists who do not work on morphology or syntax up to ...
Cited by 79 Related articles All 7 versions Cite Save More

An error-driven word-character hybrid model for joint Chinese word segmentation and POS tagging
C Kruengkrai, K Uchimoto, J Kazama, Y Wang… - … of the ACL and the 4th …, 2009 - dl.acm.org
Abstract In this paper, we present a discriminative word-character hybrid model for joint Chinese word segmentation and POS tagging. Our word-character hybrid model offers high performance since it can handle both known and unknown words. We describe our ...
Cited by 75 Related articles All 11 versions Cite Saved

Pointwise prediction for robust, adaptable Japanese morphological analysis
G Neubig, Y Nakata, S Mori - Proceedings of the 49th Annual Meeting of …, 2011 - dl.acm.org

Finding Older Papers

- Simply look at the "References" section

References

Vamshi Ambati and Alon Lavie. 2008. Improving syntax driven translation models by re-structuring divergent and non-isomorphic parse tree structures. In *Proc. AMTA*, pages 235–244.

Necip Ayan and Bonnie Dorr. 2006. Going beyond AER: an extensive analysis of word alignments and their impact on MT. In *Proc. ACL*.

David Chiang. 2007. Hierarchical phrase-based translation. *Computational Linguistics*, 33(2):201–228.

Michael Collins, Philipp Koehn, and Ivona Kucerova. 2005. Clause restructuring for statistical machine translation. In *Proc. ACL*, pages 531–540.

Yang Feng, Yang Liu, Qun Liu, and Trevor Cohn. 2012. Left-to-right tree-to-string decoding with prediction. In *Proc. EMNLP*, pages 1191–1200.

Daniel Flannery, Yusuke Miyao, Graham Neubig, and Shinsuke Mori. 2011. Training dependency parsers from partially annotated corpora. In *Proc. IJCNLP*,

Hideki Isozaki, Tsutomu Hirao, Kevin Duh, Katsuhito Sudoh, and Hajime Tsukada. 2010a. Automatic evaluation of translation quality for distant language pairs. In *Proc. EMNLP*, pages 944–952.

Hideki Isozaki, Katsuhito Sudoh, Hajime Tsukada, and Kevin Duh. 2010b. Head finalization: A simple reordering rule for SOV languages. In *Proc. WMT and MetricsMATR*.

Rasoul Samad Zadeh Kaljahi, Raphael Rubino, Johann Roturier, and Jennifer Foster. 2012. A detailed analysis of phrase-based and syntax-based machine translation: The search for systematic differences. In *Proc. AMTA*.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Proc. ACL*, pages 177–180.

Keys to Reading Lots of English

- Measure your reading speed
- Don't get stuck on one paper [1]
- Explain the papers to others [1]
- Write a summary when finished

[1] http://d.hatena.ne.jp/syou6162/20101207/1291672110

Paper Structure

3 Major Paper Structures

	Intro	Intro	Intro + Related Work
	Related Work	The Problem	The Problem
	The Problem	Proposed Method	Proposed Method
	Proposed Method	Experiments	Experiments
	Experiments	Related Work	Conclusion
	Conclusion	Conclusion	
% of Papers at ACL:	45%	35%	15%

Will explain this time

Abstract	(¼ page)
Intro	(1 page)
The Problem	(1~2 pages)
Proposed Method	(2~5 pages)
Experiments	(1~3 pages)
Related Work	(½ page)
Conclusion	(½ page)
References	(1~2 pages)

Typical Abstract

- Four sentences:

  1) What is the problem?

  2) Overview of the proposed method

  3) Merits/details of the proposed method

  4) Experimental results

Purpose of the Abstract

- Two main purposes:
  - Concisely describe the paper content
  - Decide the reviewers

ID	Title, Authors and keywords	Format	Bids
11	[Abstract]		(•) Yes ( ) Maybe ( ) No ( ) Conflict — Yes
19	[Abstract]		( ) Yes ( ) Maybe (•) No ( ) Conflict — No
26	[Abstract]		( ) Yes (•) Maybe ( ) No ( ) Conflict — Maybe

Title/Keyword/Abstract Format Want to Review?

Image: softconf.com

Example of an Abstract

- Annotation errors can significantly hurt classifier performance, yet datasets are only growing noisier […].
- In this paper, we present a robust extension of logistic regression that incorporates the possibility of mislabelling directly into the objective.
- This model can be trained through nearly the same means as logistic regression, and retains its efficiency on high-dimensional datasets.
- We conduct experiments on named entity recognition data and find that our approach can provide a significant improvement over the standard model when annotation errors are present.

[Tibshirani+ 14]

Abstract	(¼ page)
Intro	(1 page)
The Problem	(1~2 pages)
Proposed Method	(2~5 pages)
Experiments	(1~3 pages)
Related Work	(½ page)
Conclusion	(½ page)
References	(1~2 pages)

Introduction

1) Tell your story

2) Explain your contributions

That's it.

Telling your Story

- What is the problem we will solve?
- Why is the problem interesting?
- Why can't we solve it?
  (With the closest previous research?)

Explaining your Contributions

- What is your solution to the problem?
- Why is the solution exciting?

Dos and Don'ts

- Don'ts
  - “In recent years ...”
  - “The structure of this paper is …”
- Dos
  - Make the differences clear
  - Use figures
  - Ask questions
  - Make contributions clear

"In recent years ..."

- In recent years, with the spread of the Web, massive amounts of text information have become available.
- In recent years, with our increasingly international society, the need to communicate with people of other cultures is more important than ever.

"Yeah, I know..."

(Just delete it and start from the next sentence)

Make your Differences Clear

In this paper, we present a hybrid approach to sentence simplification which departs from this previous work in two main ways.

First, it combines a model encoding probabil-

...

Second, our approach is semantic based. While

...

[Narayan+ 14]

To solve above issues, this paper proposes a more general and effective framework for semi-supervised dependency parsing, referred to as *ambiguity-aware ensemble training*. Different from traditional self/co/tri-training which only use 1-best parse trees on unlabeled data, our approach adopts ambiguous labelings, represented by parse

[Li+ 14]

… departs from previous work in two ways:
First, …
Second, ...

Different from …, which only uses …, our approach can use ….

Use Figures

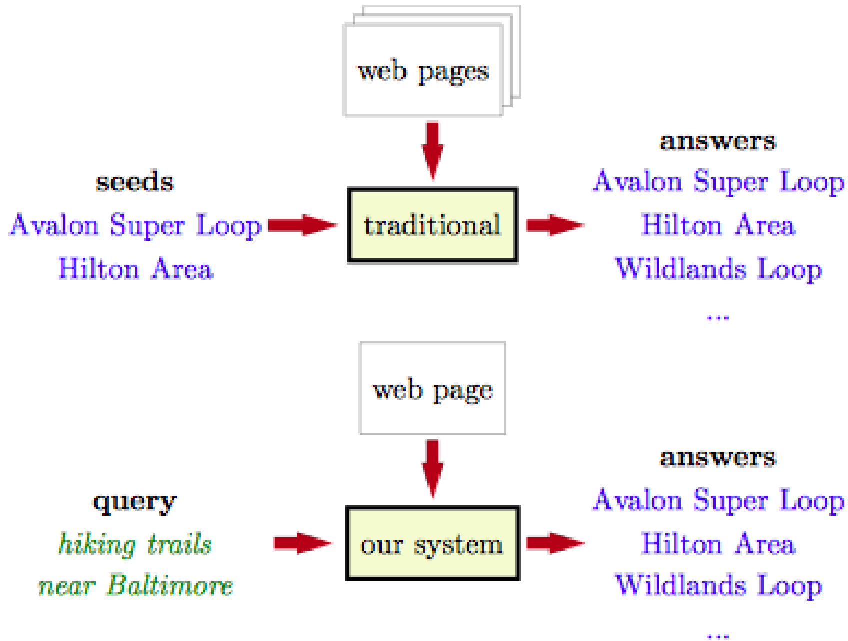


Figure 1: Entity extraction typically requires additional knowledge such as a small set of seed examples or depends on multiple web pages. In our setting, we take as input a natural language query and extract entities from a single web page.

[Pasupat+ 14]

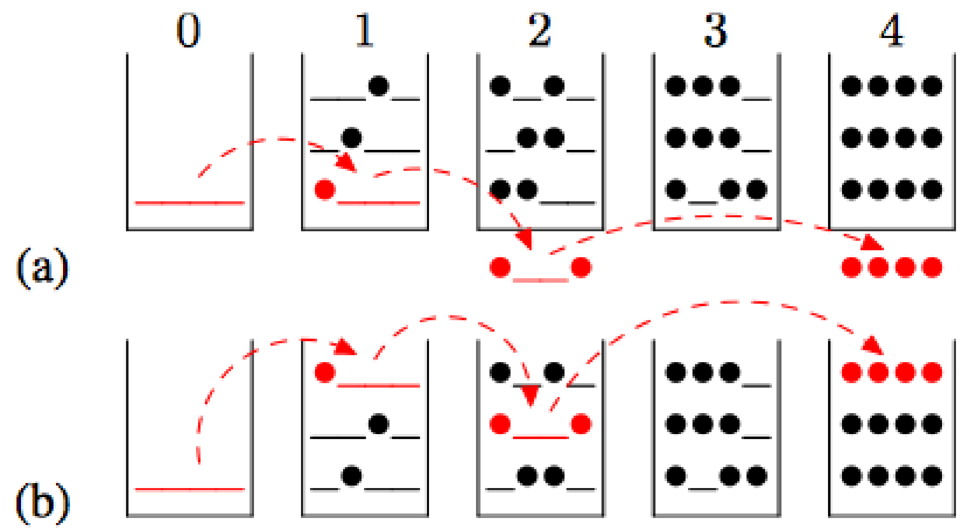


Figure 1: (a) Some potentially promising partial translations (in red) fall out of the beam (bin 2); (b) We identify such partial translations and assign them higher model scores so that they are more likely to survive the search.

[Liu+ 14]

Ask Questions

To achieve its expressiveness, CCG exhibits so-called "spurious" ambiguity, permitting many non-standard surface derivations which ease the recovery of certain dependencies, especially those arising from type-raising and composition. But this raises the question of what is the most suitable model for CCG: *should we model the derivations, the dependencies, or both?* The choice for

[Xu+ 14]

How does one make a message "successful"? This question is of interest to many entities, including political parties trying to *frame* an issue (Chong and Druckman, 2007), and individuals attempting to make a point in a group meeting. In the first

[Tan+ 14]

A question makes the reader want to know the answer!

It is also a promise of an answer.

Make your Contributions Clear

community). Below is a high-level outline of this paper.

- We formalize our approach within a probabilistic graphical model framework, inference in which yields "code-switched" text that maximizes a surrogate to the acquisition rate objective.
- We compare this global method to several baseline techniques, including the strong "high-frequency" baseline.
- We analyze the operating range in which our model is effective and motivate the near-future extension of this approach with the proposed improvements.

[Labutov+ 14]

parsing (McClosky et al., 2010). Our contributions are summarised as follows:

- Given the distribution $w_{\mathcal{S}}$ of a word w in a source domain $\mathcal{S}$, we propose a method for learning its distribution $w_{\mathcal{T}}$ in a target domain $\mathcal{T}$.
- Using the learnt distribution prediction model, we propose a method to learn a cross-domain POS tagger.
- Using the learnt distribution prediction model, we propose a method to learn a cross-domain sentiment classifier.

To our knowledge, ours is the first successful attempt to learn a model that predicts the distribution of a word across different domains.

[Bollegala+ 14]

Bullet points are effective.

"The Structure of this Paper is ..."

The structure of the paper is as follows. First, in Section 2 we introduce the formulation of the problem. In Section 3, we introduce our proposed method. In Section 4, we describe our experiments, and summarize the results. In Section 5 we describe related work, and in Section 6 we state our conclusions and discuss future work.

We can guess what it says without even reading!

Instead, References throughout the Intro

However, the utility of such a representation for summarization depends on the quality of pattern clusters. In particular, event patterns must correspond to grammatically correct sentences. Introducing an incomplete or incomprehensible pattern (e.g., *PER said PER*) may negatively affect both event detection and sentence generation. Related work on paraphrase detection and relation extraction is mostly *heuristics-based* and has relied on hand-crafted rules to collect such patterns (see Sec. 2). A standard approach is to focus on binary relations between entities and extract the dependency path between the two entities as an event representation. An obvious limitation of this approach is there is no guarantee that the extracted pattern corresponds to a grammatically correct sentence, e.g., that an essential prepositional phrase is retained like in *file for a divorce*.

In this paper we explore two novel, data-driven methods for event pattern extraction. The first, *compression-based* method uses a robust sentence compressor with an aggressive compression rate to get to the core of the sentence (Sec. 3). The second, *memory-based* method relies on a vast collection of human-written headlines and sentences to find a substructure which is known to be grammatically correct (Sec. 4). While the latter method comes closer to ensuring perfect grammaticality, it introduces a problem of efficiently searching the vast space of known well-formed patterns. Since standard iterative approaches comparing every pattern with every sentence are prohibitive here, we present a search strategy which scales well to huge collections (hundreds of millions) of sentences.

[Pighin+ 14]

Abstract	(¼ page)
Intro	(1 page)
The Problem	(1~2 pages)
Proposed Method	(2~5 pages)
Experiments	(1~3 pages)
Related Work	(½ page)
Conclusion	(½ page)
References	(1~2 pages)

No Need for Previous Research Yet

Previous research
is complicated...

Previous research
can be long...

Images:
Flickr cristiano_betta,
CollegeDegrees360

Describing your Problem

- What kind of problem? (in detail)
- Formal explanation of the problem, using variables, etc.
- Don't explain the proposed method in this section.

Explaining the Proposed Method

- Explain the intuition (most important!)
- Explain the details (secondary)

Describe the Problem with Examples!

2 Preordering for SMT

Machine translation is defined as transformation of source sentence $F = f_1 \dots f_J$ to target sentence $E = e_1 \dots e_I$. In this paper, we take the pre-ordering approach to machine translation (Xia and McCord, 2004), which performs translation as a two step process of reordering and translation (Figure 1). Reordering first deterministically transforms F into F', which contains the same words as F but is in the order of E. Translation then transforms F' into E using a method such as phrase-based SMT (Koehn et al., 2003), which can produce accurate translations when only local reordering is required.

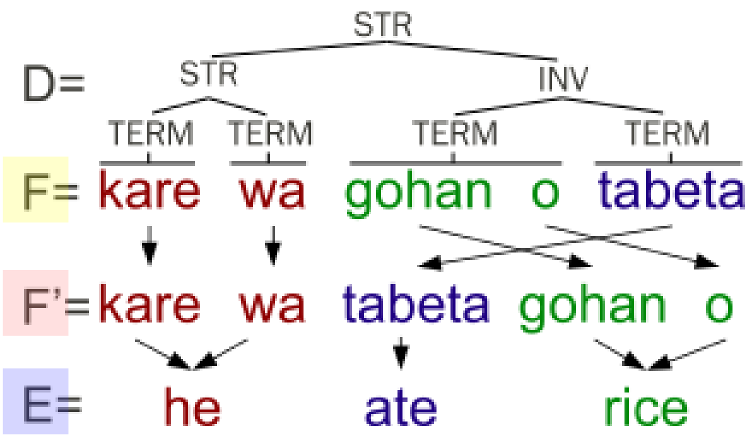


Figure 1: An example with a source sentence F reordered into target order F', and its corresponding target sentence E. D is one of the BTG derivations that can produce this ordering.

[Neubig+ 12]

Frequent Problems

- Explaining the details before the intuition
  - Details cannot be understood without intuition
- Skipping the details
  - Explain the details carefully with formulas/algorithms
- Not justifying the claims in the intro
  - The claims in the intro are a promise, fulfill them!

Abstract	(¼ page)
Intro	(1 page)
The Problem	(1~2 pages)
Proposed Method	(2~5 pages)
Experiments	(1~3 pages)
Related Work	(½ page)
Conclusion	(½ page)
References	(1~2 pages)

The Role of the Evaluation

1) Back up the claims with empirical evidence.
2) Compare other methods with the proposed method.

Many papers slack on 2)

But slacking on 2) can be dangerous!

Easy/Hard to Understand Evaluations

Easy to Understand	Hard to Understand
Evaluate on standard data (e.g. Penn Treebank, WMT)	Use your own data. Especially if not made public.
Use a standard evaluation measure (e.g. BLEU, ROUGE)	Invent your own evaluation measure.
Use recent research as a baseline and get better accuracy.	No comparison, or no statistically significant gain.

But, "Hard to Understand" does not necessarily mean "Bad."

If the research has value, do it.
(But be prepared for criticism...)

Detailed Experimental Results

- Ablation tests: Remove one feature of your method at a time and measure the accuracy decrease.

Method	Ave. prec. (%)
Proposed	**46.27**
w/o Context features	45.68
w/o Association features	45.66
w/o Semantic relation features	44.44
Base features only	41.29

Table 3: Ablation tests.

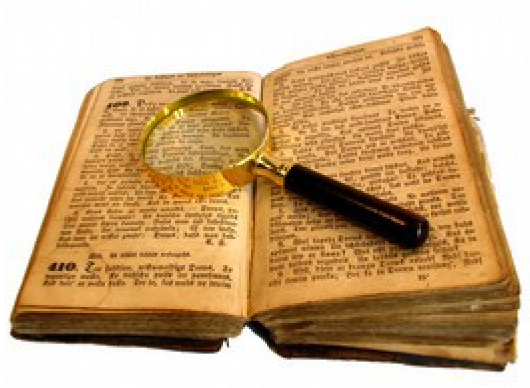
[Hashimoto+ 14]

- Examples
  - Better if you can show that examples are not flukes

Abstract	(¼ page)
Intro	(1 page)
The Problem	(1~2 pages)
Proposed Method	(2~5 pages)
Experiments	(1~3 pages)
Related Work	(½ page)
Conclusion	(½ page)
References	(1~2 pages)

Role of the Related Work Section

1) Increase readers' understanding



2) Describe this paper's differences

	A	B	C
Previous	○	×	×
Proposed	○	○	○

If a Highly Related Work is not Covered

If you didn't know:

→ Indicates incomplete understanding

If you knew:

→ Indicates intentionally hiding

Both are major problems, and can influence acceptance/rejection

No need to attack!

Attacking

Smith et al. (2015) has the serious disadvantage of not incorporating semantic context, which is known to be essential for this task.

Not Attacking

We follow in the footsteps of Smith et al. (2015), further expanding their model to allow the incorporation of not only syntactic, but also semantic information.

Dr. Smith will probably read this paper!

Abstract	(¼ page)
Intro	(1 page)
The Problem	(1~2 pages)
Proposed Method	(2~5 pages)
Experiments	(1~3 pages)
Related Work	(½ page)
Conclusion	(½ page)
References	(1~2 pages)

Conclusion

- Approx. 3 sentences about the problem, the proposed method, and the results
- Future work
    - Acknowledge incomplete parts of the work

Abstract	(¼ page)
Intro	(1 page)
The Problem	(1~2 pages)
Proposed Method	(2~5 pages)
Experiments	(1~3 pages)
Related Work	(½ page)
Conclusion	(½ page)
References	(1~2 pages)

Common Problems with References

- Reference is missing

It is important to check the references [?].

- Use of initials, full names for authors is not consistent

Philipp Koehn and Hieu Hoang. 2007. Factored translation models. In *Proc. EMNLP*.

P. Koehn, F.J. Och, and D. Marcu. 2003. Statistical phrase-based translation. In *Proc. HLT*, pages 48–54.

- Lower-case proper names

T. Kudo, H. Ichikawa, and H. Kazawa. A joint inference of deep case analysis and zero subject generation for japanese-to-english statistical machine translation. In *Proc. ACL*, pages 557–562, 2014.

- Venue is missing

P. Pasupat and P. Liang. Zero-shot entity extraction from web pages. pages 391–401, 2014.

- Venue names inconsistent

I. Labutov and H. Lipson. Generating code-switched text for lexical learning. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 562–571, 2014.

Z. Li, M. Zhang, and W. Chen. Ambiguity-aware ensemble training for semi-supervised dependency parsing. In *Proc. ACL*, pages 457–467, 2014.

Use BibTeX!

- Format of a BibTeX database:

Bracket proper names

```
@inproceedings{neubig11aclshort,
    title = {Pointwise Prediction for Robust, Adaptable {Japanese} Morphological Analysis},
    author = {Graham Neubig and Yosuke Nakata and Shinsuke Mori},
    booktitle = ACL11,
    ;;address = {Portland, USA},
    pages = {529--533},
    year = {2011}
}
```

Make venue names variables to ensure consistency and allow switching between full names/abbreviations

- Sources of BibTeX files:

Google Scholar

Pointwise prediction for robust, adaptable Japanese morphological analysis
G Neubig, Y Nakata, S Mori - Proceedings of the 49th Annual Meeting of ..., 2011 - dl.acm.org
Abstract We present a **pointwise** approach to **Japanese morphological analysis** (MA) that ignores structure information during learning and tagging. Despite the lack of structure, it is able to outperform the current state-of-the-art structured approach for **Japanese** MA, and **...**
Cited by 68 Related articles All 13 versions Cite Saved

ACL Anthology

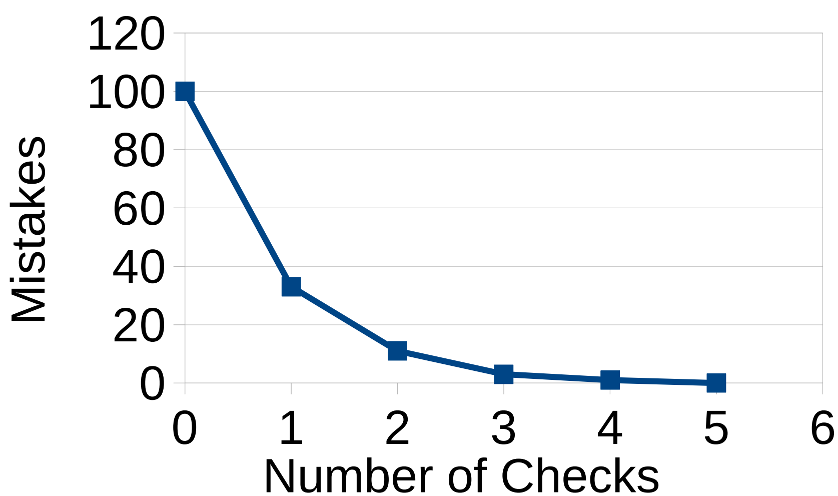
P11-2093 [bib] [software]: **Graham Neubig; Yosuke Nakata; Shinsuke Mori**
Pointwise Prediction for Robust, Adaptable Japanese Morphological Analysis

Proofreading

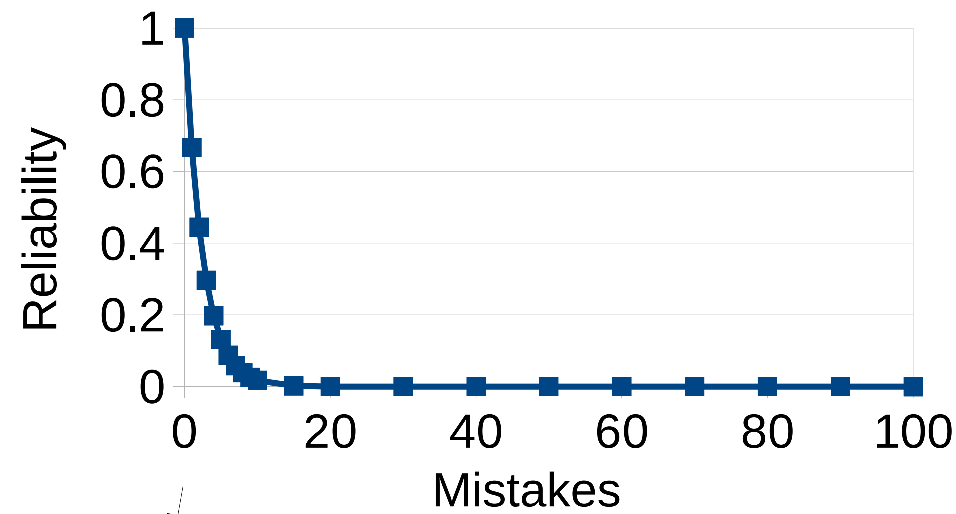
Times Proofread, Mistakes, Reliability

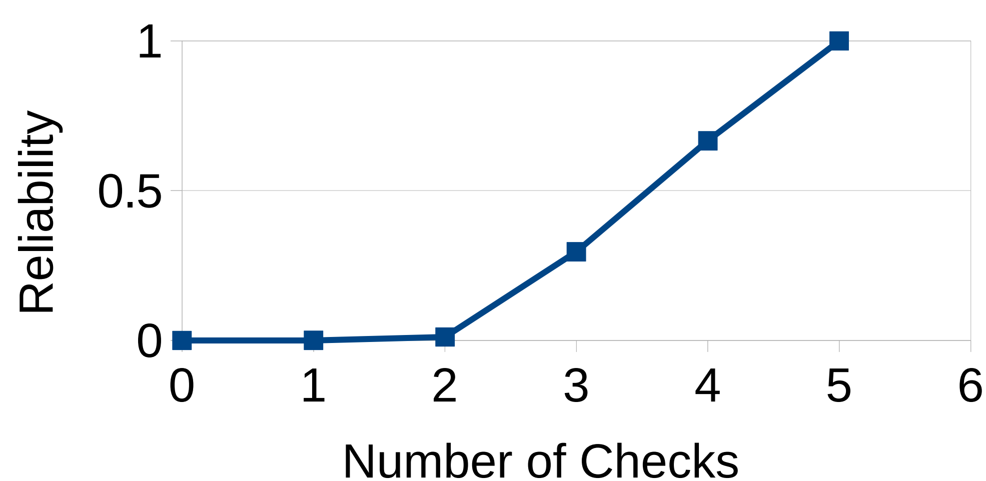
Assume: 100 mistakes in first draft

Assume: 1 proofreading reduces mistakes by 2/3



Assume: For every mistake reliability decreases 1/3





Thus, the relationship between proofreading and reliability is:

Who do we have Read the Paper?

- Ourselves
- Co-authors
- Other Researchers
- Reviewers

When we Read Ourselves

- Find mistakes, contradictions
- Find typos, formatting mistakes
- Spell check, grammar check

Print on paper and read aloud!

When we have Co-authors Read

- Essentially the same as when we read ourselves.
- Co-authors often have more experience.

Best to have a full first draft at least two weeks before the deadline.

When we have Other Researchers Read

- The opinion of someone unfamiliar with the research is essential!
- Typos and small points less important than “I didn't understand...”

Each person can only read for the first time once!

When Reviewers Read

- Decides whether the paper gets accepted
- But at the same time, often receive good advice!

The reviewers are donating their time!
Respect their advice, even if harsh.

Respecting Others' Advice

This is quite difficult...

“This person isn't reading carefully... He/she didn't understand...”
Edit so that the paper can be understood with even a quick reading.

“Comments on minor points are missing the forest for the trees...”
Sometimes minor points are important. Try to cover all your bases.

English Basics for Papers

The Very Basics

- Use spell-check
- Space punctuation correctly
  - Need a space before: ([{
  - Need a space after:)] } : . , ! ?
- Be careful of capitals
  - In the actual content, only capitalize person/place names. Methods should not be capitalized.
  - In titles, capitalize content words, not function words.

Customary Expressions for Papers

- a lot → many
- means → indicates
- really → very
- But, → However,
- Also, → In addition,
- So, → Thus,

Articles (the/a)

- Don't use
  - Proper nouns such as person names
  - Words that describe actions, ones ending in “...ion”
  - Plural, unless it specifies a particular set of things
- Use
  - Basically everything else. Don't forget.
- If you're not sure, check in a dictionary whether the noun is “countable” or not.

Use Active Verbs

Passive	Active
A corpus was gathered and a model was trained.	We gathered a corpus and trained a model.
It can be seen that ...	We can see that ...
In the next section, … is described.	The next section describes ….
It may be thought that this will ….	You may think that this will ...

After your Paper is Accepted

Paper != Influence

“It is easy to become a leader in a field. Just make a new field. It's much more difficult to find followers.”

Another Anonymous Professor

Why is a Paper Influential?

- Content:
  Presents or solves an important problem
- Presentation:
  Publicize your work at conferences, online, etc.
- Ease of Use:
  Provision of tools/data

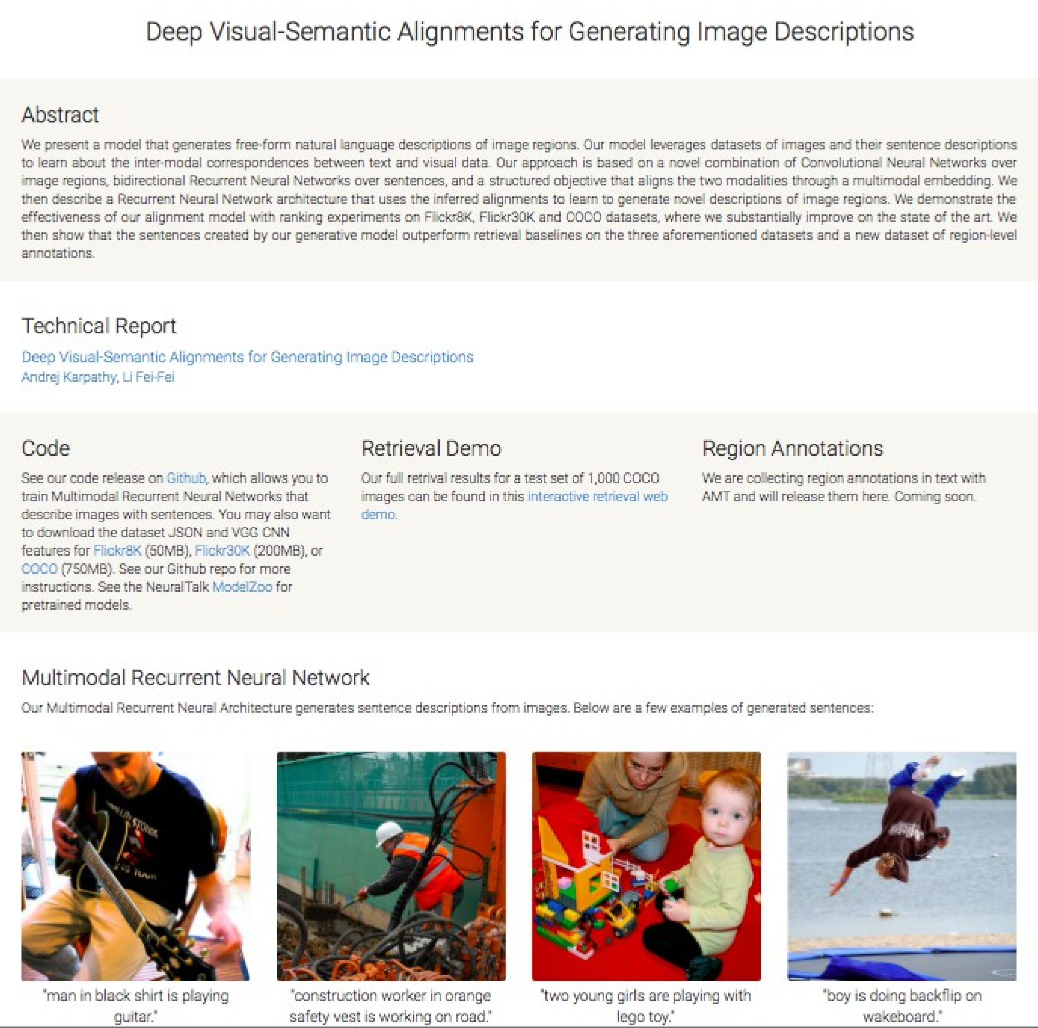
Elements of a Good Presentation

- Reduce the amount of text in slides (Prepare a script if necessary)
- Put effort into the first several slides (Like your intro, tell your story)
- Lots of practice (If it's your first presentation, 50 times is not too much.)

Releasing your Code/Data

	Code	Data
Released	Others can test on their own data set.	Easy to replicate the results.
Not Released	Takes time to re-implement. Not sure if the details are right.	Re-creating data is difficult to impossible.

Project Pages



Deep Visual-Semantic Alignments for Generating Image Descriptions

Abstract

We present a model that generates free-form natural language descriptions of image regions. Our model leverages datasets of images and their sentence descriptions to learn about the inter-modal correspondences between text and visual data. Our approach is based on a novel combination of Convolutional Neural Networks over image regions, bidirectional Recurrent Neural Networks over sentences, and a structured objective that aligns the two modalities through a multimodal embedding. We then describe a Recurrent Neural Network architecture that uses the inferred alignments to learn to generate novel descriptions of image regions. We demonstrate the effectiveness of our alignment model with ranking experiments on Flickr8K, Flickr30K and COCO datasets, where we substantially improve on the state of the art. We then show that the sentences created by our generative model outperform retrieval baselines on the three aforementioned datasets and a new dataset of region-level annotations.

Technical Report

Deep Visual-Semantic Alignments for Generating Image Descriptions
Andrej Karpathy, Li Fei-Fei

Code

See our code release on Github, which allows you to train Multimodal Recurrent Neural Networks that describe images with sentences. You may also want to download the dataset JSON and VGG CNN features for Flickr8K (50MB), Flickr30K (200MB), or COCO (750MB). See our Github repo for more instructions. See the NeuralTalk ModelZoo for pretrained models.

Retrieval Demo

Our full retrival results for a test set of 1,000 COCO images can be found in this interactive retrieval web demo.

Region Annotations

We are collecting region annotations in text with AMT and will release them here. Coming soon.

Multimodal Recurrent Neural Network

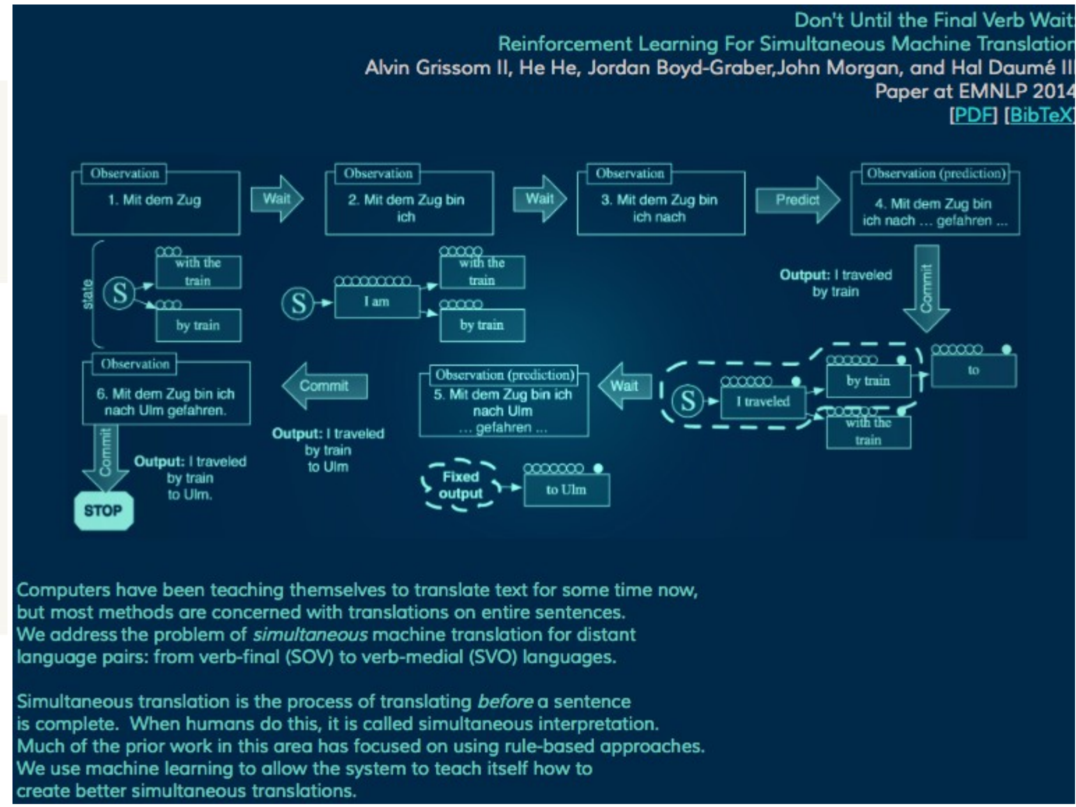
Our Multimodal Recurrent Neural Architecture generates sentence descriptions from images. Below are a few examples of generated sentences:

"man in black shirt is playing guitar."

"construction worker in orange safety vest is working on road."

"two young girls are playing with lego toy."

"boy is doing backflip on wakeboard."



Don't Until the Final Verb Wait:
Reinforcement Learning For Simultaneous Machine Translation
Alvin Grissom II, He He, Jordan Boyd-Graber, John Morgan, and Hal Daumé III
Paper at EMNLP 2014
[PDF] [BibTeX]

Computers have been teaching themselves to translate text for some time now, but most methods are concerned with translations on entire sentences. We address the problem of *simultaneous* machine translation for distant language pairs: from verb-final (SOV) to verb-medial (SVO) languages.

Simultaneous translation is the process of translating *before* a sentence is complete. When humans do this, it is called simultaneous interpretation. Much of the prior work in this area has focused on using rule-based approaches. We use machine learning to allow the system to teach itself how to create better simultaneous translations.

[Karpathy 14]

[Grissom 14]

Conclusion

Take-home Messages

- Go for it!
- Tell your story
- Use figures/examples
- Have many people read your paper

References

Advice About Paper Writing

- Simon Peyton Jones: How to Write a Great Research Paper
  http://research.microsoft.com/en-us/um/people/simonpj/papers/giving-a-talk/writing-a-paper-slides.pdf
- Graham Neubig: Paper style guide
  http://phontron.com/paper-guide.php

Paper Examples

- D. Bollegala, D. Weir, and J. Carroll. Learning to predict distributions of words across domains. In Proc. ACL, pages 613–623, 2014.
- C. Hashimoto, K. Torisawa, J. Kloetzer, M. Sano, I. Varga, J.-H. Oh, and Y. Kidawara. Toward future scenario generation: Extracting event causality exploiting semantic relation, context, and association features. In Proc. ACL, pages 987–997, 2014.
- I. Labutov and H. Lipson. Generating code-switched text for lexical learning. In Proc. ACL, pages 562–571, 2014.
- Z. Li, M. Zhang, and W. Chen. Ambiguity-aware ensemble training for semi-supervised dependency parsing. In Proc. ACL, pages 457–467, 2014.
- L. Liu and L. Huang. Search-aware tuning for machine translation. In Proc. EMNLP, pages 1942–1952, 2014.
- S. Narayan and C. Gardent. Hybrid simplification using deep semantics and machine translation. In Proc. ACL, pages 435–445, 2014.
- P. Pasupat and P. Liang. Zero-shot entity extraction from web pages. In Proc. ACL, pages 391–401, 2014.
- C. Tan, L. Lee, and B. Pang. The effect of wording on message propagation: Topic- and author-controlled natural experiments on twitter. In Proc. ACL, pages 175–185, 2014.
- J. Tibshirani and C. D. Manning. Robust logistic regression using shift parameters. In Proc. ACL, pages 124–129, 2014.
- W. Xu, S. Clark, and Y. Zhang. Shift-reduce CCG parsing with a dependency model. In Proc. ACL, pages 218–227, 2014.

Project Page Examples

- Andrej Karpathy: Deep Visual-Semantic Alignments for Generating Image Descriptions
  http://cs.stanford.edu/people/karpathy/deepimagesent/
- Alvin Grissom II: Don't Until the Final Verb Wait: Reinforcement Learning For Simultaneous Machine Translation
  http://www.umiacs.umd.edu/~alvin/research/simtrans/