Tel(s)-Telle(s) Signs:
Highly Accurate Automatic Cross-Lingual Hypernym Discovery

Ada Wan
ada.wan@uzh.ch
adawan919@gmail.com

1. What is a hypernym?

Citrus fruits such as oranges, lemons or grapefruits contain a lot of vitamin C.
HYPERNYM   HYPONYMS

citrus fruits
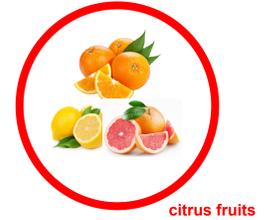
2. What is the big deal?

- uncover ontological structures (e.g. for knowledge base development, information extraction, natural language understanding)
- "Hearst patterns": lexico-syntactic patterns for hypernym and hyponym extraction by Marti Hearst (1992), e.g.:
  NP_0 *such as* {NP_1, NP_2 ..., (*and|or*)} NP_n,
  where it is implied that for all NP_i, $1 \leq i \leq n$, hyponym(NP_i, NP_0); i.e. hypernym(NP_0, NP_i).

3. What's so hard?

i. Urban birds in cities such as pigeons, crows, ...
ii. Urban birds in cities such as Paris, Rome, ...
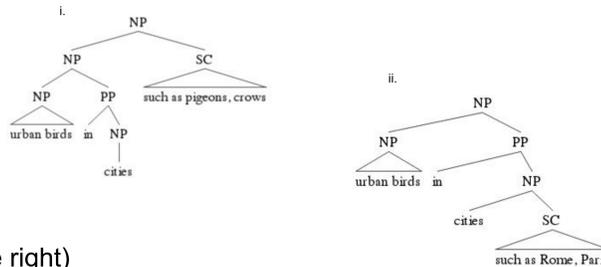
Which of the two NPs — *urban birds* or *cities* — should be the hypernym?

Possible approaches:
- pure surface form: always select the closest noun (*BUT* can only get one right)
- syntactic: select the head of the NP as the hypernym (*BUT* presupposes correct syntactic analysis)
  probabilistic parsers, selecting the one that has a higher likelihood (*BUT* we can do better than this!)
- morphological? (*BUT* English is not too morphologically endowed...)

French Grammar Review:
Adjectives and nouns inflect according to gender and number. They can be:
- masculine singular,
  e.g. *tel, livre*
- feminine singular,
  e.g. *telle, référence*
- masculine plural,
  e.g. *tels, livres*
- feminine plural,
  e.g. *telles, références*.

4. A crosslingual approach:

À l'examen, vous pourrez utiliser des **livres** de référence **tels** que grammaires et encyclopédies.
(English translation: At the examination, you can use reference books (lit.: books of reference) such as encyclopedias and grammars.)
tels agrees with *livres*, but not with *référence* in gender and number. *tels* and *livres* are masculine plural, while *référence* is feminine singular.

This means that we can actually use the inflectional morphological cues inherent in the French language to detect the correct hypernym!
tel- que is our *such as*, the ending in *tel-* tells us the morphological information that the hypernym should have.

5. It works for French, how well can it work for English?

Our experiment:

Data:
- two sets of French-English **parallel corpora** (used: the EMEA corpus made from PDF documents form the European MEdicines Agency (Tiedemann, 2009) and the news commentary training data from the WMT 2014 shared task)

Tools:
- a **morphological analyzer** (used: *morfette* (Chrupala et al. 2008))
- **word alignment** tool (used: *GIZA++* (Och and Ney, 2003) with the grow-diag-final-and heuristic (Koehn et al. 2007))

Methods:
- tokenize, filter sentences with length over 100 for the EMEA corpus and 200 for the WMT corpus (the former can be noisier).
- tag French texts using morphological analyzer above.
- align French-English parallel texts.
- examine 100 unique sentence pairs (first 50 from each corpus) manually.

Evaluate:

A. is the closest noun preceding *tel- que* that agrees with *tel-* in gender & number the hypernym in sentence?
B. did *morfette* predict the correct outcome in question A above?
C. is the hypernym in French aligned with the hypernym in English?

6. Results:

	total number of unique sentences evaluated	hypothesis correct	French hypernym extractable (using *morfette*)	English hypernym extractable (using *GIZA++*)	
				best	worst
EMEA	50	50 (100%)	45 (90%)	41.5 / 44 (94.32%)	41.5 / 50 (83%)
WMT14 news commentary	50	47 (94%)	46 (92%)	46 (92%)	

Percentages in parentheses next to raw count. *best* indicates a total count that excludes sentences that are so misaligned that the relevant pattern is not available in English at all, *worst* is when these cases are included in the total.

7. Sample sentence pairs:

French:
Informer votre médecin le plus rapidement possible si vous avez des **signes** d'insuffisance cardiaque **tels qu'**une dyspnée inhabituelle ou une augmentation rapide de poids, huperhydratation cellulaire ou œdèmes.

English:
Inform your doctor as soon as possible if you experience **signs** of heart failure **such as** unusual shortness of breath or rapid increase in weight or localised swelling (oedema).
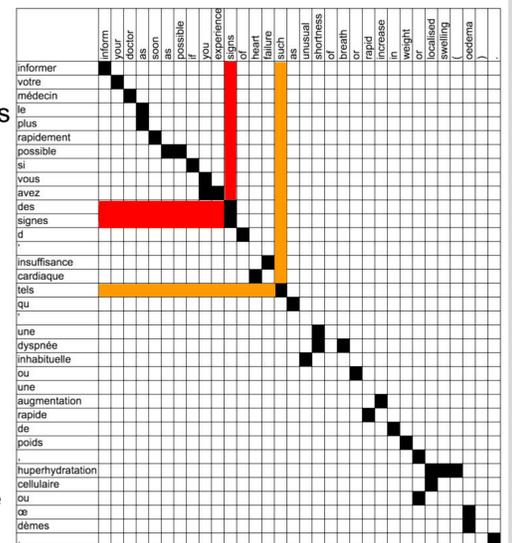
French:
APTIVUS, co-administré avec le ritonavir à faible dose, doit être utilisé avec précaution chez les patients pouvant présenter un risque accru de saignement en raison d'un traumatisme, d'une chirurgie ou d'antécédents médicaux autres, ou chez ceux recevant des **traitements** connus pour augmenter le risque de saignement **tels que** les antiagrégants plaquettaires et les anticoagulants, ou chez ceux qui prennent de la vitamine E.

English:
APTIVUS, co-administered with low dose ritonavir, should be used with caution in patients who may be at risk of increased bleeding from trauma, surgery or other medical conditions, or who are receiving **medicinal products** known to increase the risk of bleeding **such as** antiplatelet agents and anticoagulants or who are taking supplemental vitamin E.

Selected references:
- Wauter Bosma and Piek Vossen. 2010. Bootstrapping language neutral term extraction. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, editors, *Proceedings of the International Conference on Language Resources and Evaluation*, LREC 2010, 17-23 May 2010, Valletta, Malta. European Language Resources Association.
- Ruiji Fu, Bing Qin, and Ting Liu. 2013. Exploiting multiple sources for open-domain hypernym discovery. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, EMNLP 2013, 18-21 October 2013, Grand Hyatt Seattle, Seattle, Washington, USA, A meeting of SIGDAT, a Special Interest Group of the ACL, pages 1224–1234. ACL.
- Marti A. Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *14th International Conference on Computational Linguistics*, COLING 1992, Nantes, France, August 23-28, 1992, pages 539–545.
- Els Lefever, Marjan Van de Kauter, and Véronique Hoste. 2014. Evaluation of automatic hypernym extraction from technical corpora in english and dutch. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asunción Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation* (LREC-2014), Reykjavik, Iceland, May 26-31, 2014., pages 490–497. European Language Resources Association (ELRA).
- Emmanuel Morin and Christian Jacquemin. 2004. Automatic acquisition and expansion of hypernym links. *Computers and the Humanities*, 38(4):363–396.
- Alan Ritter, Stephen Soderland, and Oren Etzioni. 2009. What is this, anyway: Automatic hypernym discovery. In *Learning by Reading and Learning to Read*, Papers from the 2009 AAAI Spring Symposium, Technical Report SS-09-07, Stanford, California, USA, March 23-25, 2009, pages 88–93. AAAI.