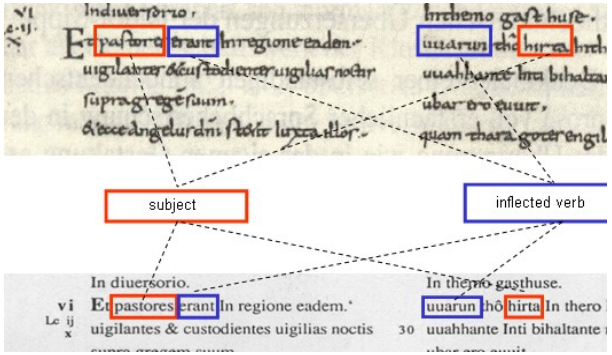
Multilingual and Parallel Corpora
Conclusion

Amir Zeldes

amir.zeldes@georgetown.edu

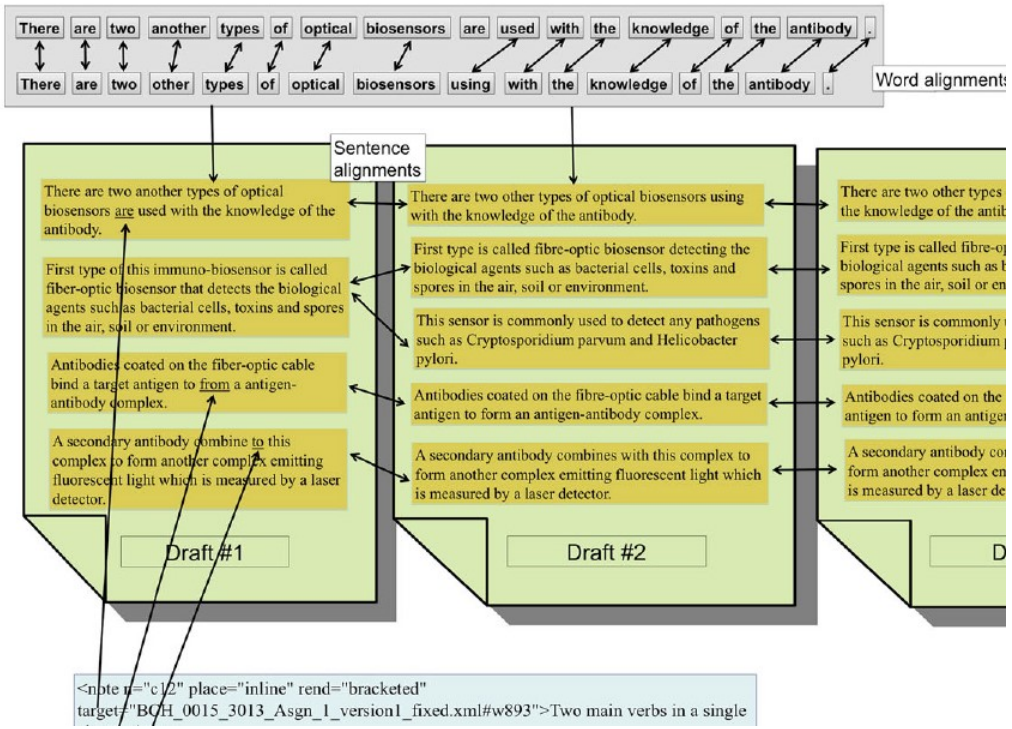
So what is a parallel?



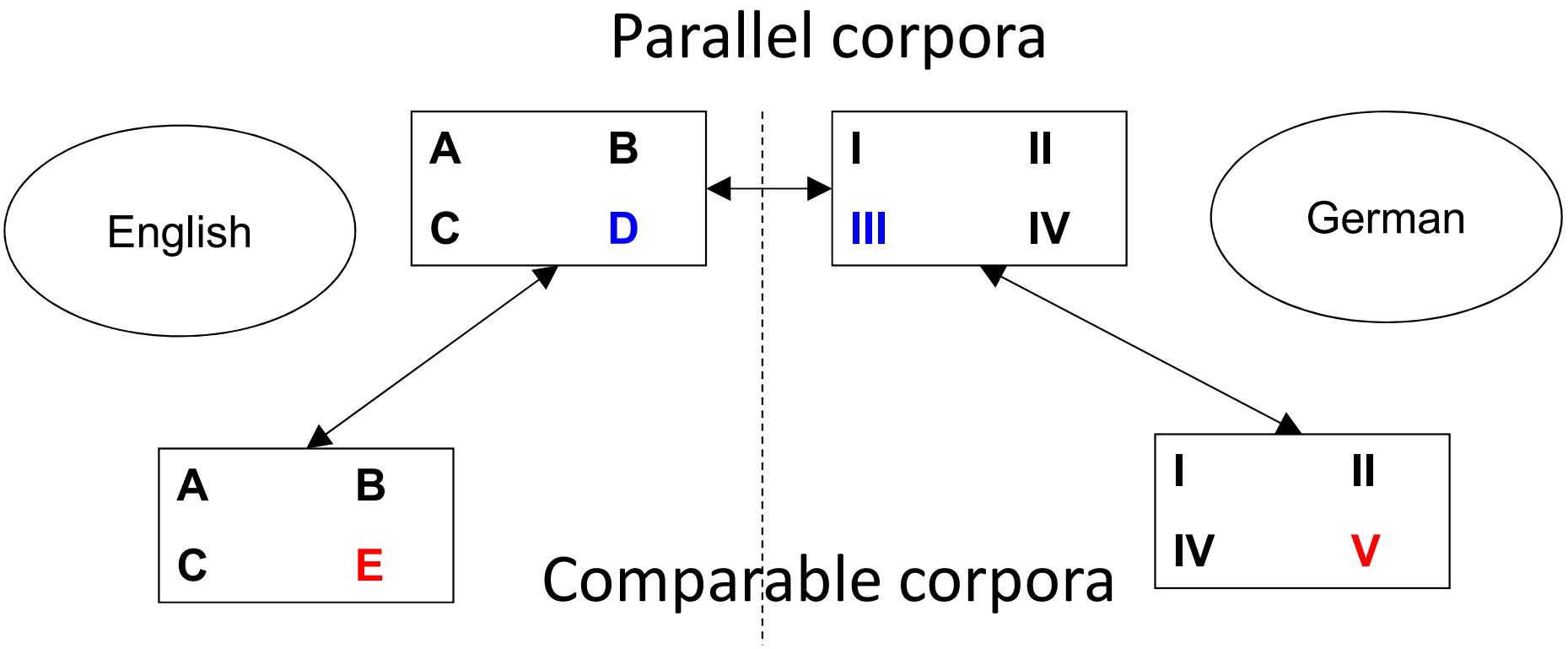
Besa
ⲙⲡⲣ ⲕⲧⲉ ⲡ ⲣⲁⲡ ⲉ ⲩ ⲭⲟⲗⲏ ⲁⲩⲱ ⲡ ⲕⲁⲣⲡⲟⲥ ⲛ ⲧ ⲇⲓⲕⲁⲓⲟⲥⲩⲛⲏ ⲛ ⲟⲩ ⲥⲓⲱⲉ
Shenoute XF
ⲉ ⲩ ⲡⲱⲱⲛⲉ ⲙ ⲡ ⲣⲁⲡ ⲉ ⲩ ⲭⲟⲗⲏ ⲁⲩⲱ ⲡ ⲕⲁⲣⲡⲟⲥ ⲛ ⲧ ⲇⲓⲕⲁⲓⲟⲥⲩⲛⲏ ⲉ ⲩ ⲥⲓⲱⲉ ⲉ ⲩ ⲱⲡ ⲙ ⲡ ⲕⲁⲕⲉ ⲛ ⲟ
ⲟⲩⲟⲉⲓⲛ ⲛ ⲕⲁⲕⲉ ⲉ ⲩ ⲭⲱ ⲁⲉ ⲟⲛ ⲙⲙⲟ ⲥ ⲉ ⲡ ⲉⲧ ⲥⲁⲱⲉ ⲭⲉ ⲏ ⲣⲟⲩⲟ ⲁⲩⲱ ⲡ ⲉⲧ ⲣⲟⲩⲟ ⲭⲉ ⲏ ⲥⲁⲱⲉ

Amos 6:12

für	neuen	Aufgaben	
für	neue	Aufgaben	
	CHA		
für	neue	Aufgaben	



Authenticity and direction



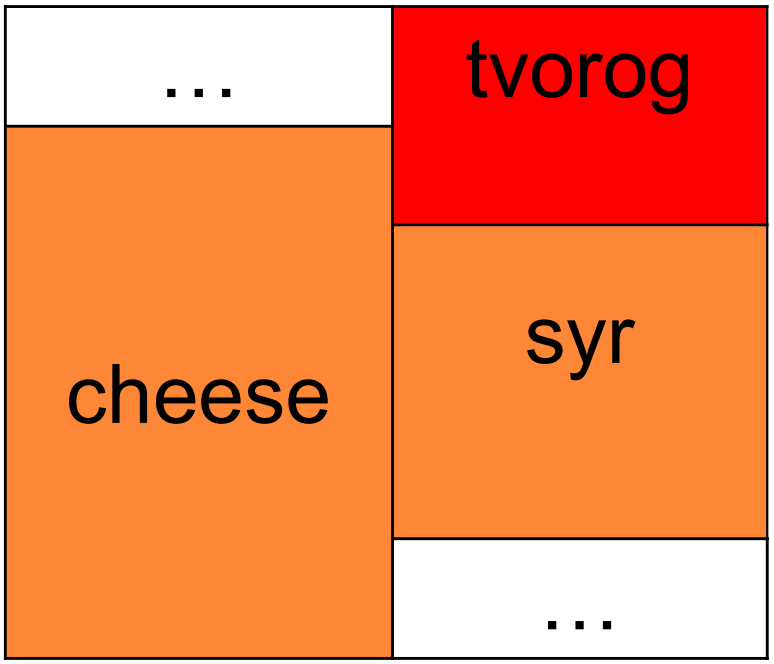
Translation

Many critics, no defenders,
Translators have but two regrets;
When they "hit" no-one remembers,
When they "miss" no-one forgets.

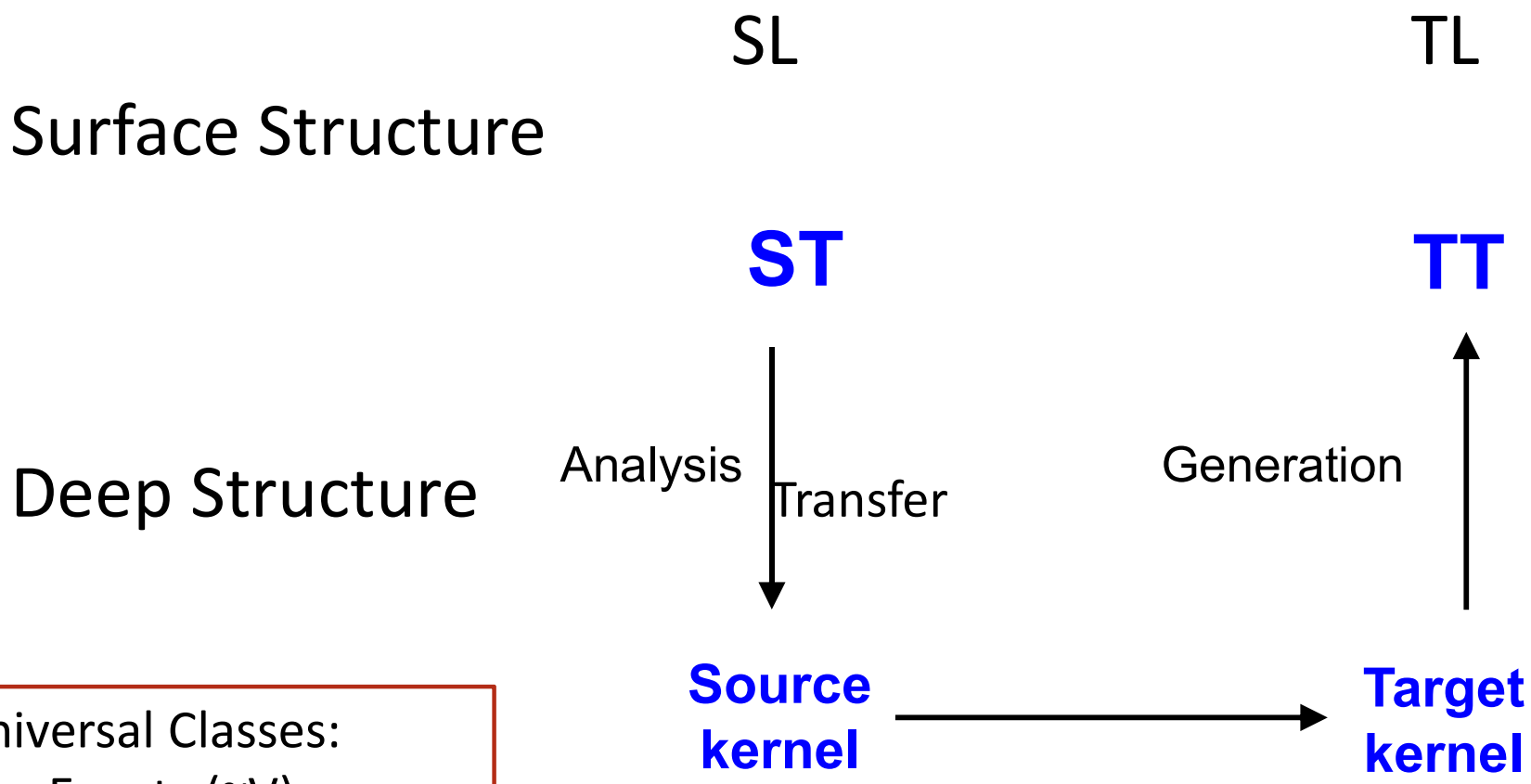
Wilson (2009), Translators on Translating

- What are the basic operations of creating a parallel?
  - Theory?
  - Typology?
- What makes a translation good?

Meaning equivalence



Nida & Taber's kernel approach



Universal Classes:
- Events (~V)
- Objects (~N)
- Properties (~A)
- Relations (~P)

Koller's multilevel equivalences

1. **Denotative** equivalence – same extension in the worlds
2. **Connotative** equivalence – corresponding choice of words, same stylistic connotations
3. **Text-normative** equivalence – same text type/register, corresponding conventions
4. **Pragmatic** equivalence – same effect on recipient, perceived the same
5. **Formal** equivalence – the same formal relationships between linguistic elements, including word games and linguistic style

Catford's shifts

- **Level shifts** – different levels (linguistic systems)
- **Category shifts** - same level
- Shift types:
  - Structural Shift
    I like **it** — **es** gefällt **mir** (it pleases me)
  - Class Shift (POS)
    Verner's Law — das **Vernersche** Gesetz (adj)
  - Unit/Rank Shift
    es ist **un**möglich — it's **not** possible (**im**possible)
  - Intra-system Shift
    er hat **das** Bein gebrochen — he broke **his** leg (explicit possesor)

Functional equivalence (Toury)

- Content is determined by function; literal translation is irrelevant:

Alarm Signal	**Notbremse**
To stop train	Griff nur bei
pull handle	Gefahr ziehen
Penalty £50	Jeder Mißbrauch
for improper use	wird bestraft

Good translation

- Goals?
  - Genre, purpose
  - Intended readership
- Metrics
  - Subjective
  - Association guidelines (IoL, ATA,ITI)
  - Computational metrics
    - BLEU
    - ROUGE
    - METEOR

Translation Universals

- Processes suggested by Baker (1996):
  - Explicitation
  - Simplification
  - Normalization (a.k.a. ‘levelling out’)

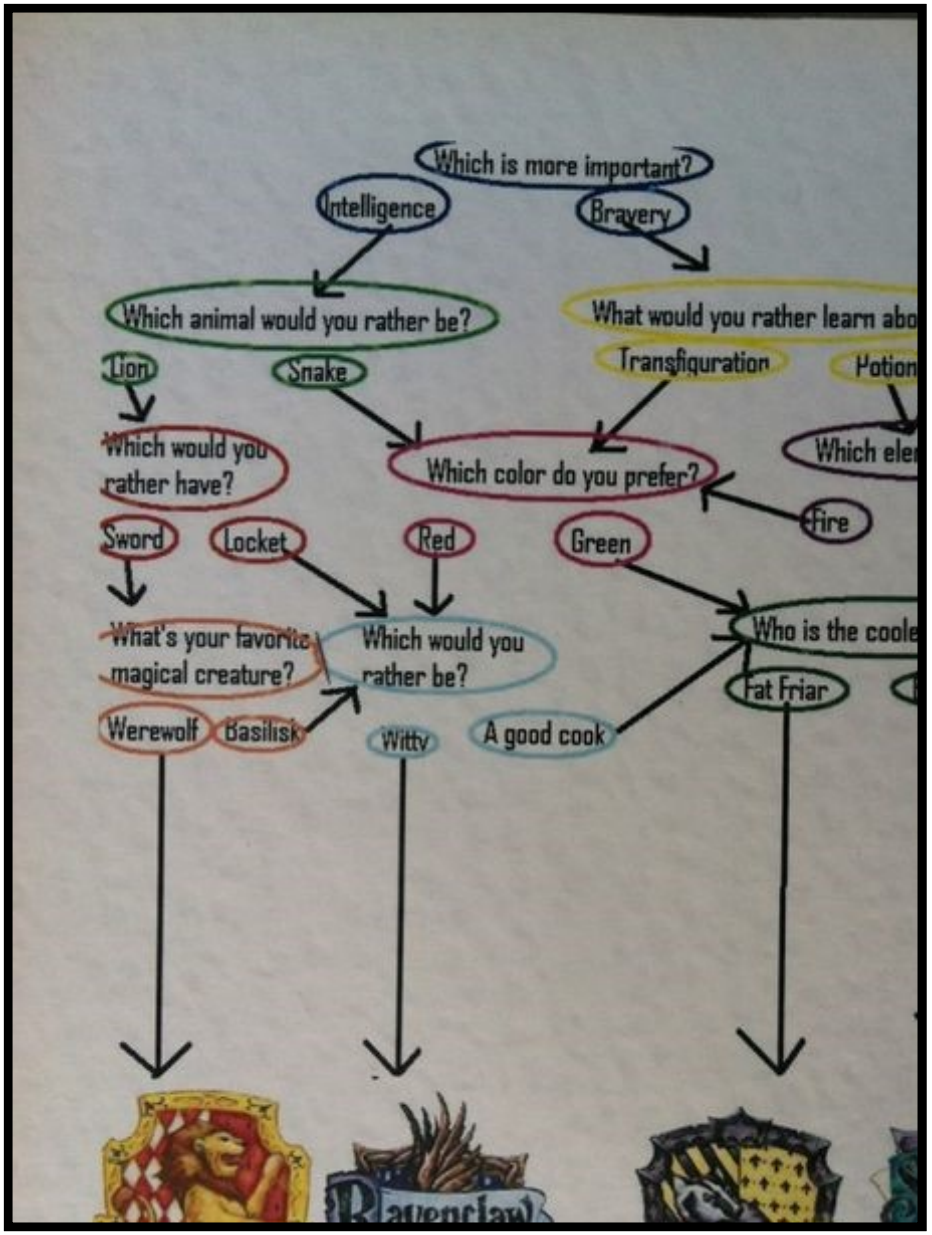
Explicitation

And now there were only three people left to be sorted.

Jetzt waren nur noch drei Schüler übrig , deren Haus bestimmt werden mußte .

Now were only still three pupils remaining, whose house determined become had-to

- sorted = into houses
- People = Schüler



Simplification

*but **there was no escaping** Dudley ' s gang , who visited the house every single day*

*doch Dudleys Bande , die das Haus Tag für Tag heimsuchte , **konnte er nicht entkommen***

Yet Dudley's gang, which the house day for day plagued could he not escape

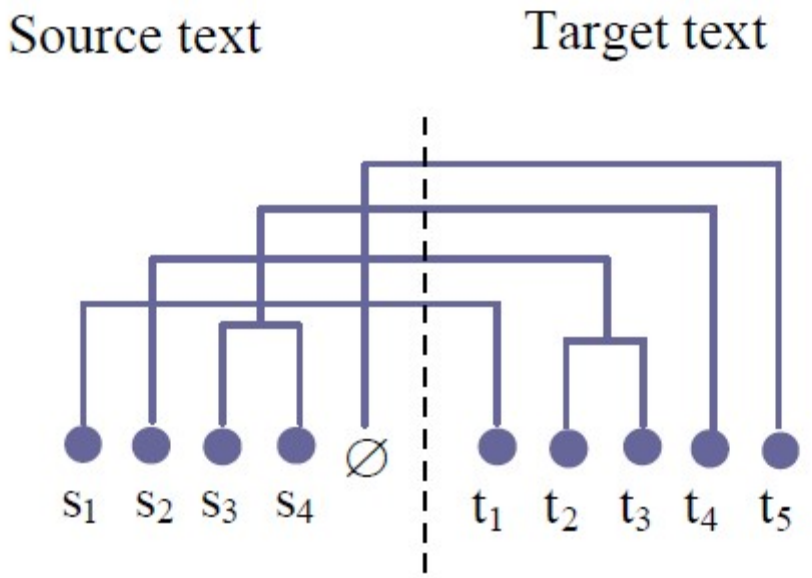
- "Standard" sentences and simple constructions are preferred

Normalization / Levelling

- How can non-standard language be translated?
  - *Begbie was hard, but not so hard that he didn't shite it off twenty years in Saughton.*
  - *Y por duro que Begbie fuese , veinte años en prision no los iba a aguantar .*
  - *Begbie était dur , mais pas au point de se foutre de vingt ans de taule .*

Sentence alignment

- Full / partial
- 1:1, 1:2, 2:1 …
- Reordering
- Null alignment
- Hierarchical alignment:



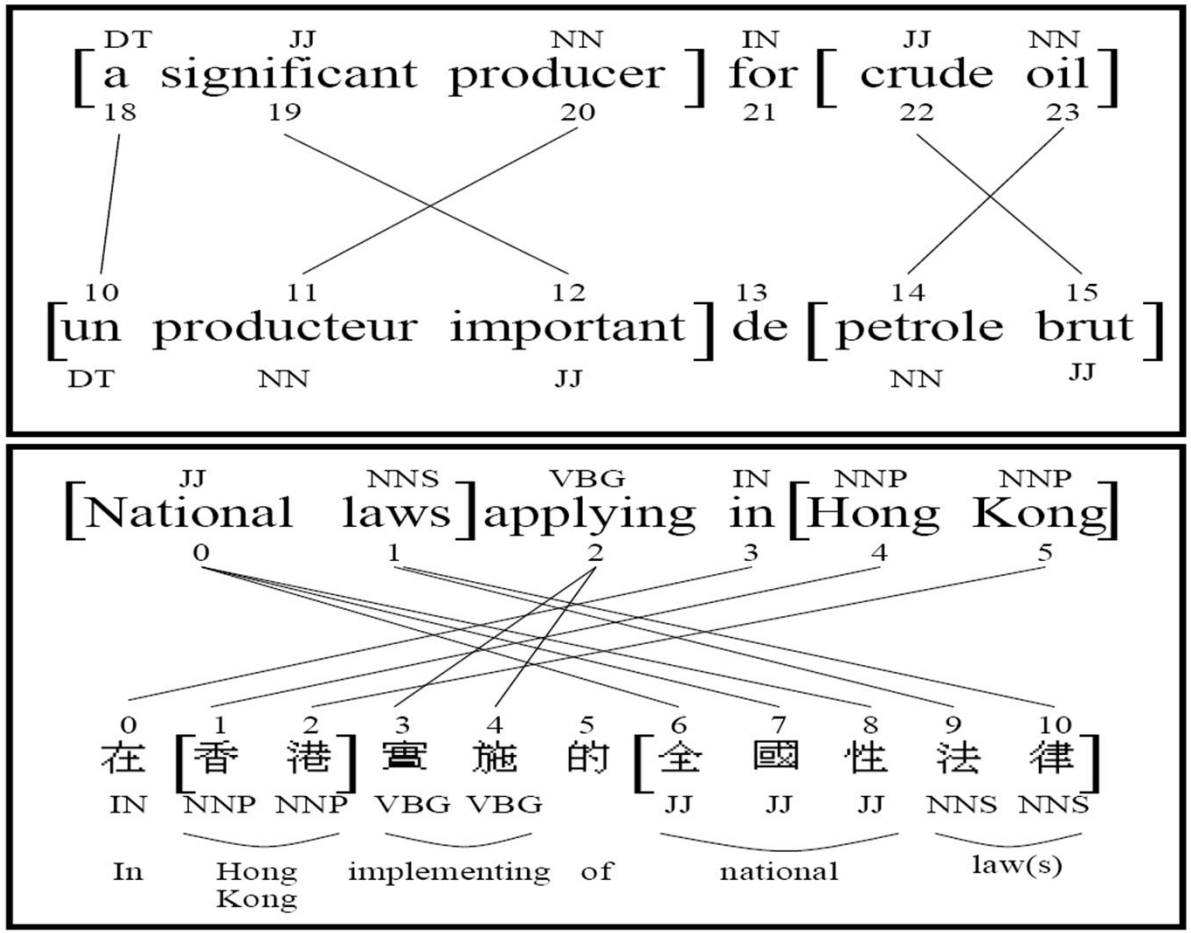
Automatic approaches

- In the meantime , I should like to observe a minute ' s silence .
- You will be aware from the press and television that there have been a number of bomb explosions and killings in Sri Lanka .
- One of the people assassinated very recently in Sri Lanka was Mr Kumar Ponnambalam , who had visited the European Parliament just a few months ago .

- Σας καλώ να σηκωθείτε για αυτή την ενός λεπτού σιγή .
- Κυρία Πρόεδρε , επί ενός θέματος διαδικασίας .
- Θα έχετε ενημερωθεί από τον τύπο και την τηλεόραση ότι συνέβησαν ορισμένες εκρήξεις βομβών και φόνοι στη Σρι Λάνκα .
- Ένας από τους ανθρώπους που δολοφονήθηκαν πολύ πρόσφατα στη Σρι Λάνκα ήταν ο κ. Kumar Ponnambalam , ο οποίος είχε επισκεφθεί το Ευρωπαϊκό Κοινοβούλιο μόλις πριν λίγους μήνες

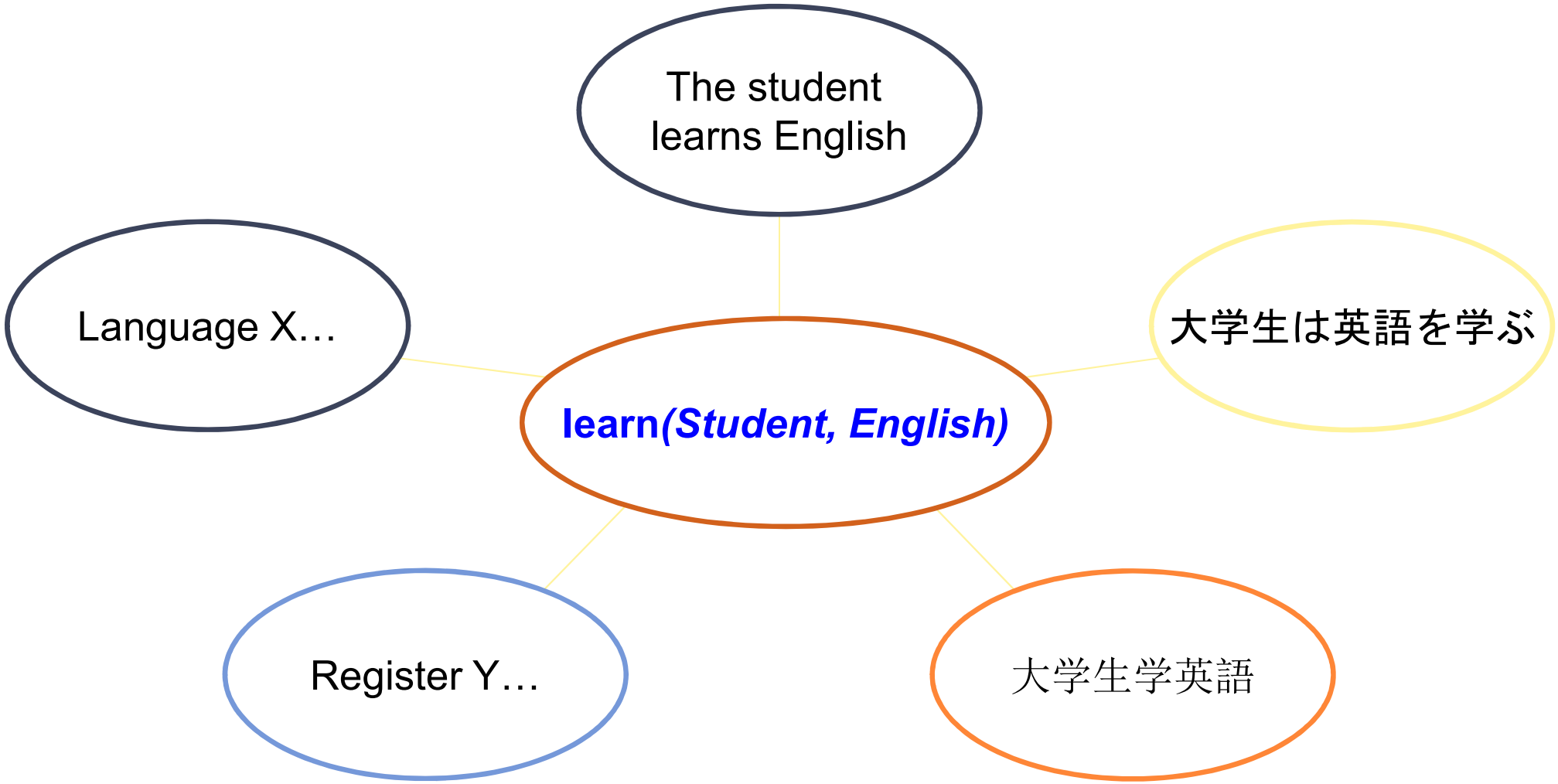
- Length
- Anchors (and how to get them…)

Word alignment and projection

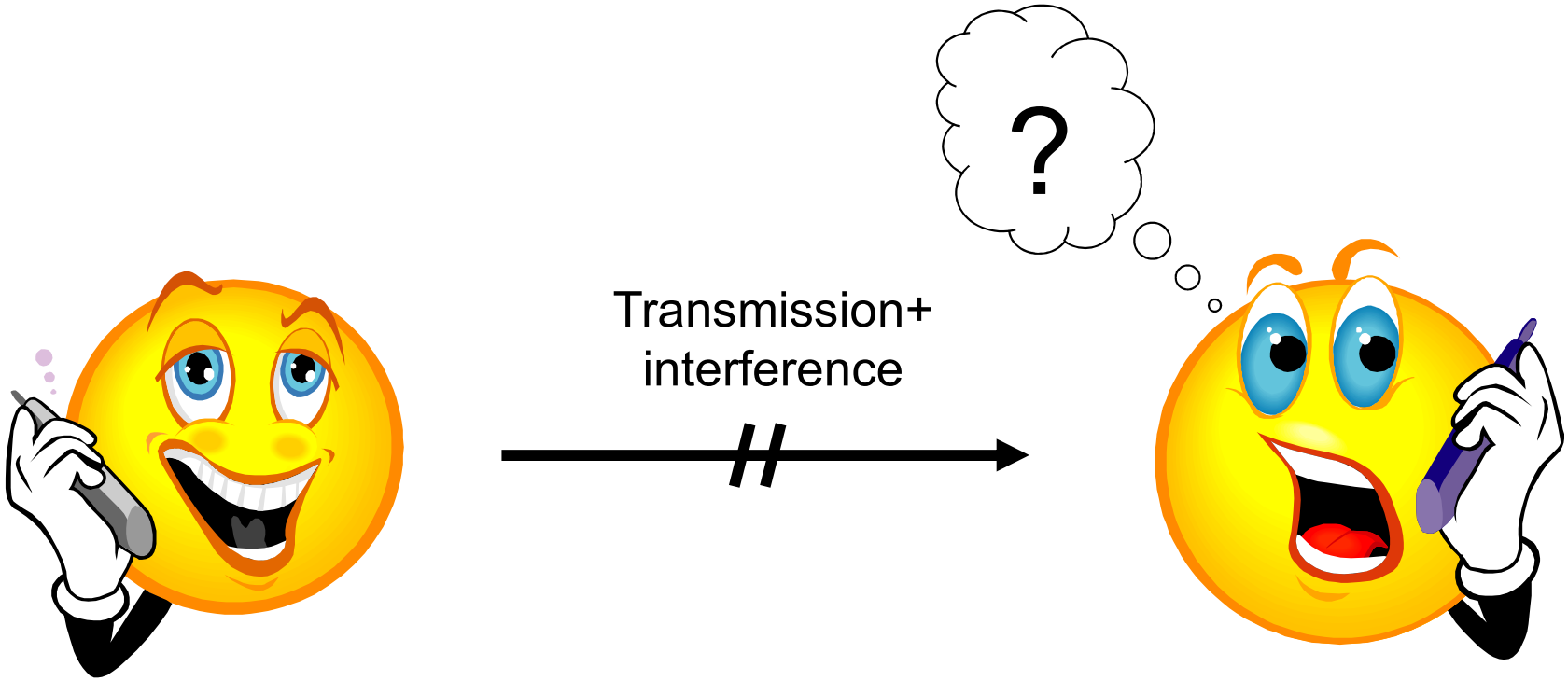


[Yarowsky/Ngai 2001]

MT – interlingua vs. direct rules



MT – Noisy Channel Model

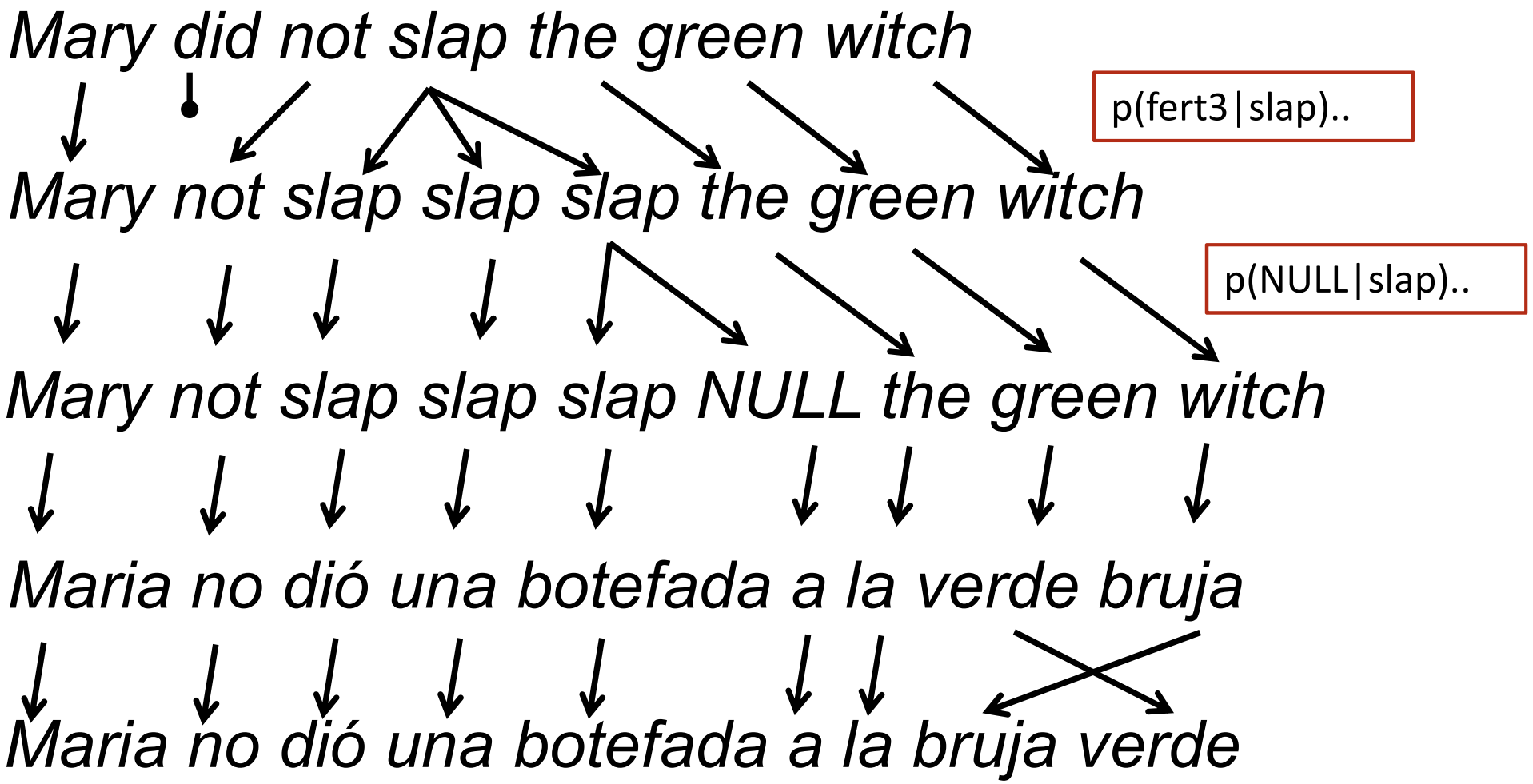


- I'm running a little late
- I-… ru-… … li-…l late

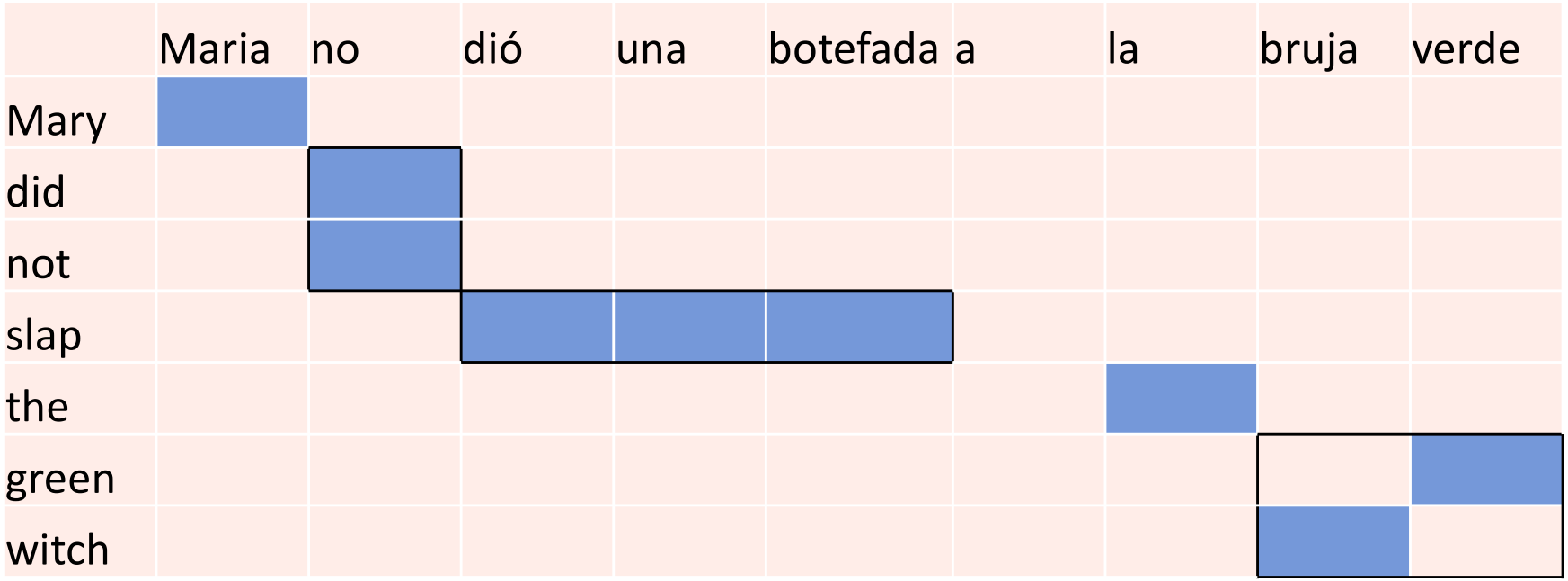
Are stats good enough?

Clients do not sell pharmaceuticals in Europe => **Clientes no venden medicinas en Europa**	
1a. Garcia and associates .	1b. Garcia y asociados .
2a. Carlos Garcia has three associates .	2b. Carlos Garcia tiene tres asociados .
3a. his associates are not strong .	3b. sus asociados no son fuertes .
4a. Garcia has a company also .	4b. Garcia tambien tiene una empresa .
5a. its clients are angry .	5b. sus clientes estan enfadados .
6a. the associates are also angry .	6b. los asociados tambien estan enfadados .
7a. the clients and the associates are enemies .	7b. los clients y los asociados son enemigos .
8a. the company has three groups .	8b. la empresa tiene tres grupos .
9a. its groups are in Europe .	9b. sus grupos estan en Europa .
10a. the modern groups sell strong pharmaceuticals.	10b. los grupos modernos venden medicinas fuertes.
11a. the groups do not sell zenzanine .	11b. los grupos no venden zanzanina .
12a. the small groups are not modern .	12b. los grupos pequenos no son modernos .

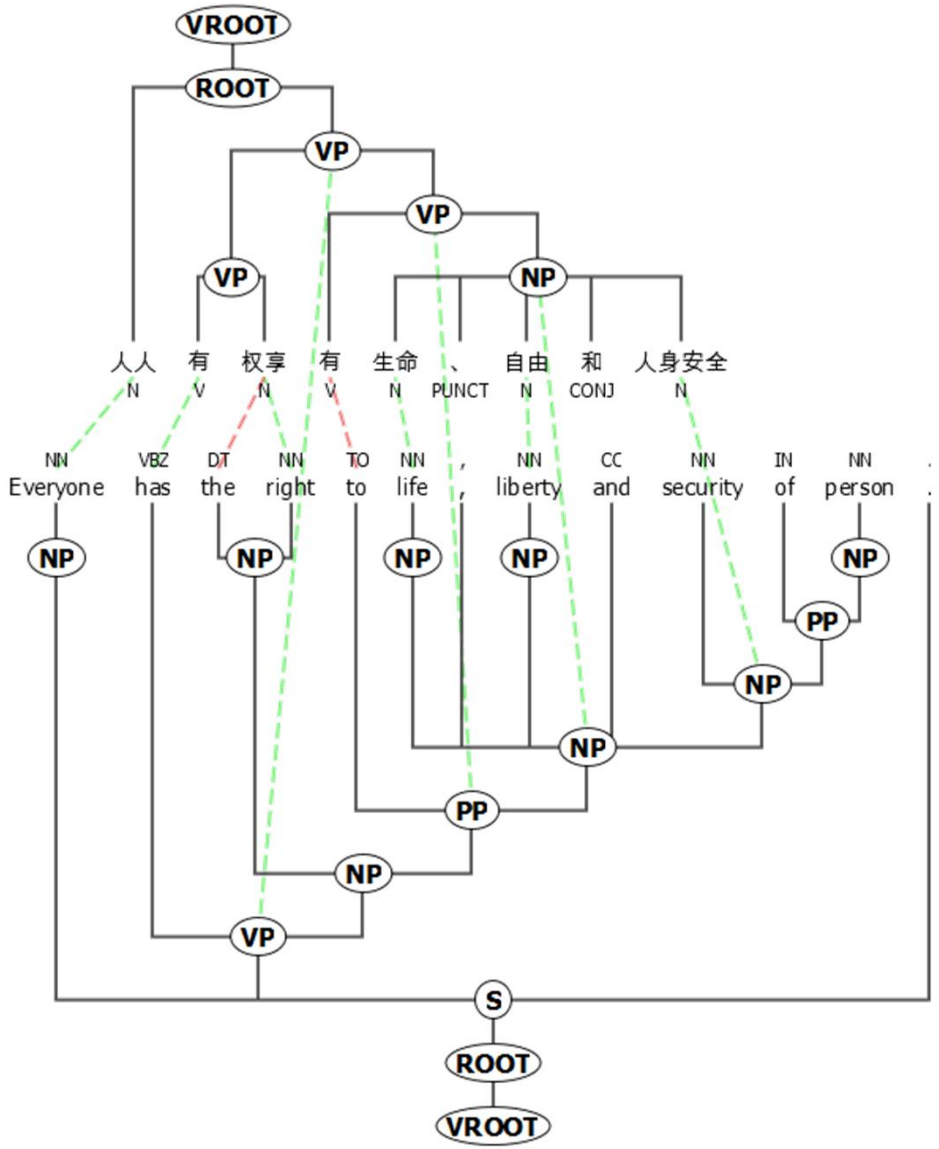
SMT – IBM Model 3



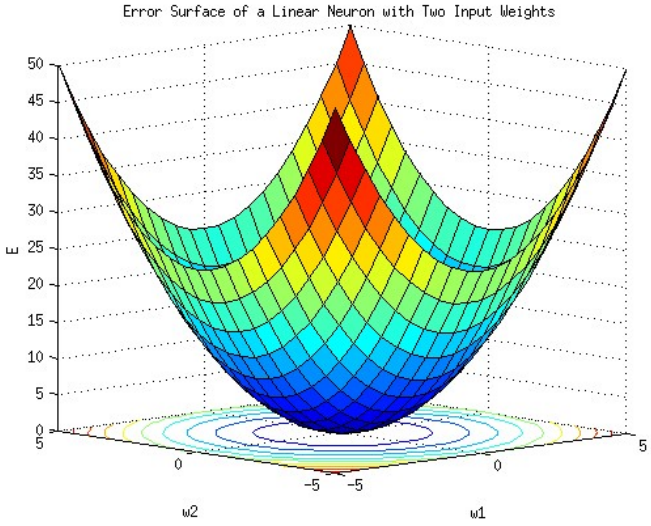
MT – phrase alignment

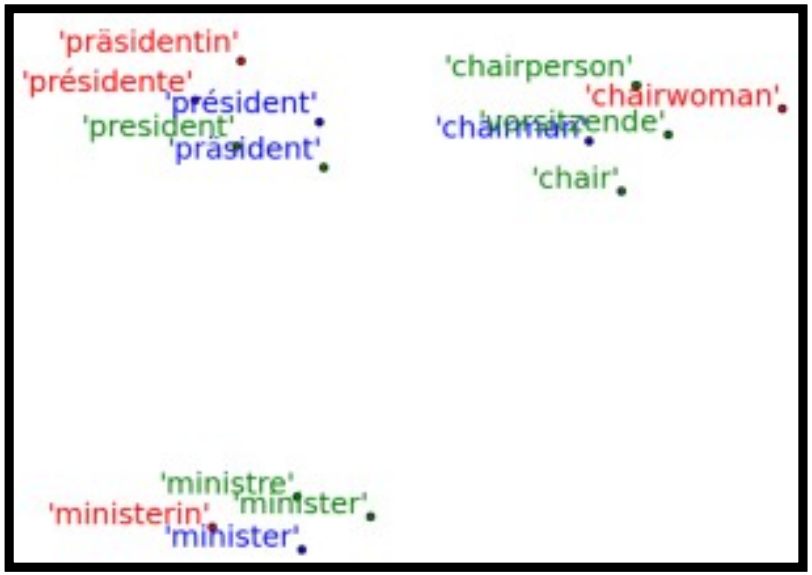


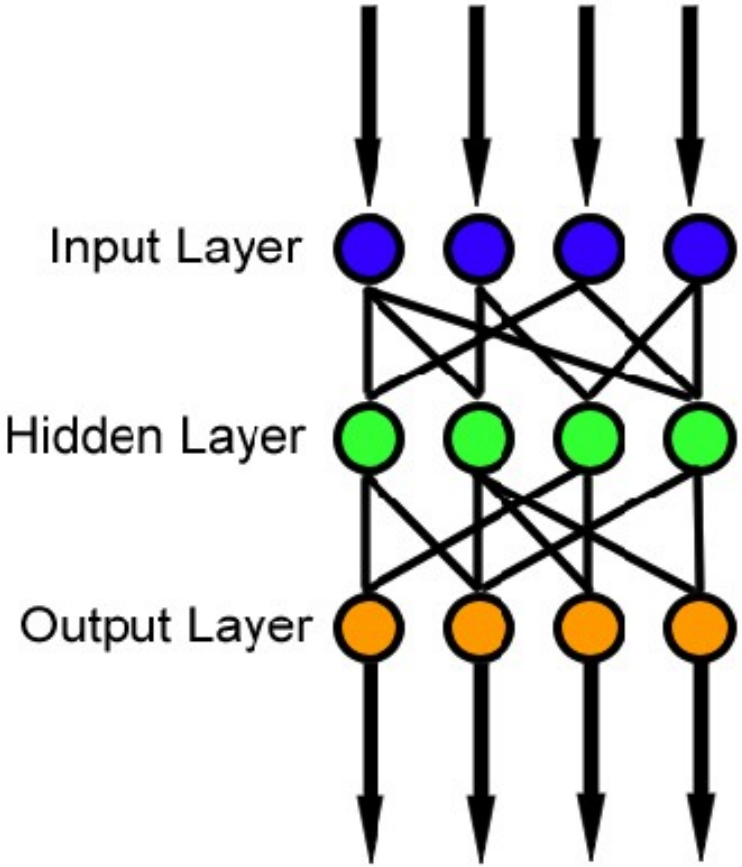
Contrast - TreeAligner



Neural approaches







And beyond…

- Historical data
- Learner data
  - Target hypotheses
  - Revisions
- Partial alignment
  - Wikipedia (non-)translations
  - Questions and answers??
  - What else?