

Translation

A New **?** Frontier

Martin Kay

Stanford University and
The University of the Saarland

Martin Kay

Machine Translation

1

The European Union



Martin Kay

Machine Translation

2

The European Union

The cost of translation, interpretation, publication, and legal services involved in making Irish an official EU language is estimated at just under €3.5 million a year.

English
Finnish
French
German
Greek
Italian
Portuguese
Spanish
Swedish

23 languages
2,500 (12.5%) of 20,000 staff
1% of the annual budget
40% of administration costs.

Irish
Latvian
Lithuanian
Maltese
Polish
Romanian
Slovene
Slovak

$$\binom{23}{2} = 253$$

Martin Kay

Machine Translation

3

CATERPILLAR

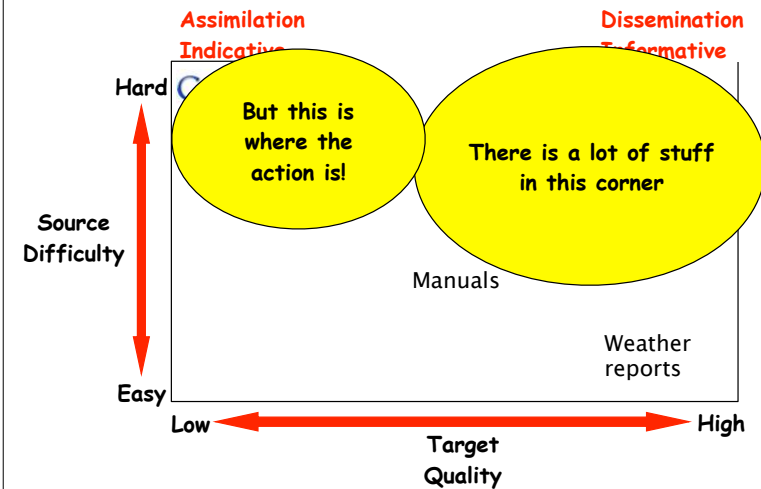
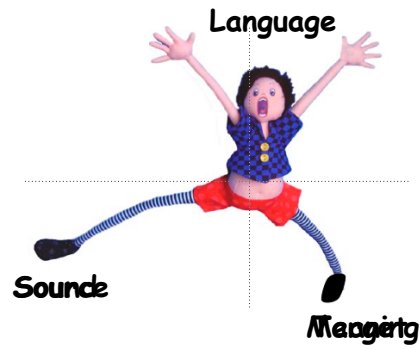
300 authors and
illustrators
800 English pages per
day
Translation into 14
languages

Maintenance Manuals
Operation and Troubleshooting
Guides
Disassembly and Specifications
Manuals
Assembly Manuals
Testing and Special Instructions
Adjustment Guides
Systems Operation Bulletins

Martin Kay

Machine Translation

4



What is Translation?

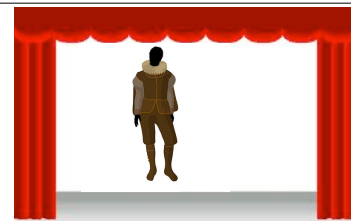
- A text that is based on a text in another language and which
 - has the same meaning
 - conveys the same information
 - has the same effect on its readers
 - gives the gist of the original
 - explains the original
- It depends what you want
- Generally a mixture of several of these

Generally Speaking—

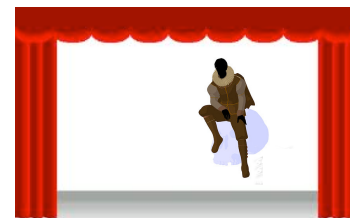
Foreign reader's experience should be as close as possible to that of a domestic reader. They should pass through parallel mental states.

No mention of meaning!

Reference



He sat down



Situation 1

Language

Situation 2

Update semantics

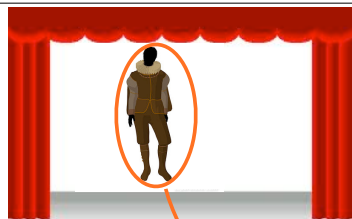


Dateline:
Damascus
Athens
Moscow

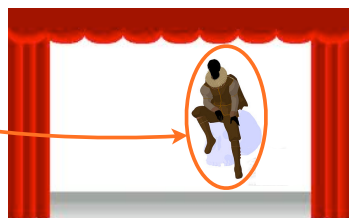


Situation Updater





Problem 1:
Keep track of the **things**



Reference



Blood from a Stone

13

There are a lot of things

She took his wallet.

- I saw it in her hand
- I saw her do it
- I could not believe it



Translation

14

but only a few things —

- are on stage
- are in the spotlight
- preserve the momentum of what is being said.



Translation

15

She took his wallet.

- I saw it in her hand
the thing
- I saw her do it
that
the deed
- I could not believe it
what she did

Martin Kay

Machine Translation

16

She took his wallet.

- I saw it
 - It was in her hand
 - She was holding it
- I saw her do it
 - I saw her take it
 - She took it as I watched
 - She did it before my very eyes
- I could not believe it
 - I did not believe she could do such a thing

She took his wallet.

These do not mean the same thing out of context
They do mean the same thing in context

- I could not believe she could do it
- It was unbelievable (to me)

The sense of a word tells you what properties a thing must have for that word to refer to it.

Things can be referred to by words with different senses.

Maybe choose the one that gives the most fluent result

Types have senses
Tokens have referents

Syntactic Processing

- Find words with similar senses
- Put them in a fluent order
- Translation in Language teaching
 - Popular perception of translation
 - Machine translation

Syntactic Processing

- Find words with similar senses
- Put them in a fluent order
- St. Jerome: Vulgate (4th Century)
 - 20th Century
 - Translator is anonymous

- Words often have several senses.
- You often cannot find words with similar senses.
- Even when you can, they often cannot be ordered fluently.

Trouble is

Different grammatical structures give different meanings.

The dog saw the cat
The cat saw the dog

Der Hund sah die Katze
Den Hund sah die Katze



Versichern Sie Sich daß Sie nichts in dem Zug vergeßen haben



Assurez-vous que vous n'avez rien oublié dans le train



Martin Kay

Machine Translation

25



Versichern Sie Sich daß Sie nichts in dem Zug vergeßen haben



Assurez-vous que vous n'avez rien oublié dans le train

Martin Kay

Machine Translation

26



Versichern Sie Sich daß Sie nichts in dem Zug vergeßen haben



Assurez-vous que vous n'avez rien oublié dans le train

Martin Kay

Machine Translation

27



Versichern Sie Sich daß Sie nichts in dem Zug vergeßen haben



Assurez-vous que vous n'avez rien oublié dans le train

Martin Kay

Machine Translation

28



Versichern Sie Sich daß **Sie** nichts in
dem Zug vergeßen haben



Assurez-vous que **vous** n'avez rien
oublié dans le train



Versichern Sie Sich daß Sie **nichts** in
dem Zug vergeßen haben



Assurez-vous que vous **n'**avez **rien**
oublié dans le train



Versichern Sie Sich daß Sie nichts **in**
dem Zug vergeßen haben



Assurez-vous que vous n'avez rien
oublié **dans** le train



Versichern Sie Sich daß Sie nichts in
dem Zug vergeßen haben



Assurez-vous que vous n'avez rien
oublié dans **le** train



Versichern Sie Sich daß Sie nichts in dem Zug vergeßen haben



Assurez-vous que vous n'avez rien oublié dans le train



Versichern Sie Sich daß Sie nichts in dem Zug vergeßen haben



Assurez-vous que vous n'avez rien oublié dans le train



Versichern Sie Sich daß Sie nichts in dem Zug vergeßen haben



Assurez-vous que vous n'avez rien oublié dans le train



Versichern Sie Sich daß Sie nichts in dem Zug vergeßen haben



Assurez-vous que vous n'avez rien oublié dans le train



Make sure that you have forgotten nothing on the train

Make sure that you have left nothing on the train



Versichern Sie Sich daß Sie nichts in dem Zug vergeßen haben



Assurez-vous que vous n'avez rien oublié dans le train



Make sure that you take all your belongings with you

Semantic Processing

Replace words and phrases by words and phrases that mean the same

Reorder

Pragmatic Processing

Replace words and phrases by words and phrases that will effect the same changes in the scene

Pragmatic translation often involves addition or removal of substantive information from what is explicitly expressed.

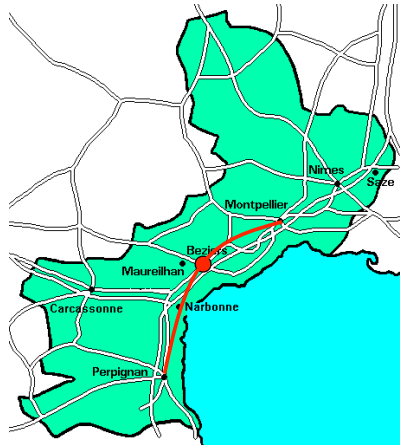


Does this train go to
Perpignan?

No, it stops in Beziers.

Fährt dieser Zug nach
Perpignan?

Nein, er { hält
endet } in
Béziers



Martin Kay

Translation

41

Est-ce que c'est ta cousine?

Non, je n'ai pas de cousine.



female
Is that your ~~male~~ cousin?

female
No. I don't have a ~~male~~ cousin.

Is that woman your cousin?

Martin Kay

Machine Translation

42

Est-ce que c'est ta cousine?

Non, je n'ai pas de cousine.



Is that your cousin?

No. I don't have a cousin.

girl
Is that ~~woman~~ your cousin?

Martin Kay

Machine Translation

43

Abstraction

Martin Kay

Machine Translation

44

Structural Linguistics

Linguistic Chemistry

Elements

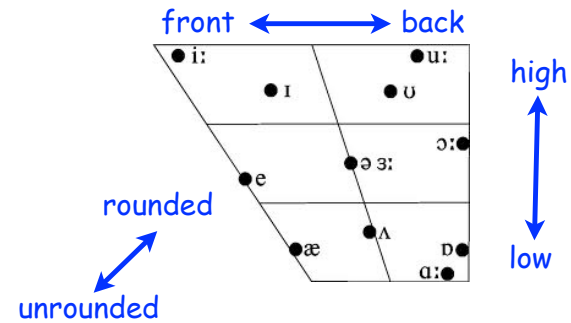
Phoneme
Morpheme
Sememe (?)
Phrase
(Distinctive) feature



Putting Linguistics back into Computational Linguistics

45

Dimensions



Putting Linguistics back into Computational Linguistics

46

Phonology

North American Flap

$\left[\begin{array}{l} + \text{stop} \\ + \text{consonant} \\ + \text{alveolar} \end{array} \right] \rightarrow [r] / \left[\begin{array}{l} + \text{vowel} \\ + \text{stressed} \end{array} \right] \text{ ____ } \left[\begin{array}{l} + \text{vowel} \\ - \text{stressed} \end{array} \right]$

lætər → lærər



Putting Linguistics back into Computational Linguistics

47

Generalization

One statement covers many instances.



Putting Linguistics back into Computational Linguistics

48

Morphographemics

```

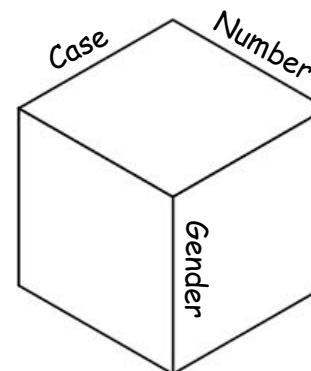
define sib [j | s | x | z | sh | ch] ;
define consonant [b | c | d | f | g | h | j | k | l | m | n | p |
                  q | r | s | t | v | w | x | y | z] ;
define vowel [a | e | i | o | u] ;
define boundary [.#. | %+];
define optional [ %? (->) 0] ;
define YtoIE [ y -> i e || consonant _ EM alpha] ;
define ItoY [ i e -> y ||
define Edeletion1 [ e -> 0 || vowel _ EM ] ;
define Edeletion2 [ e EM e -> EM
define Einsertion [ [..] -> e ||
define gemination [ b -> b b, c -> c k, d -> d d, f -> f f, g -> g g,
                    l -> l l, m -> m m, n -> n n, p -> p p, r -> r r,
                    s -> s s, t -> t t || vowel _ EM vowel ] ;
define DiacriticDeletion [ diacritic -> 0 ] ;
define BoundaryDeletion [ [BM | EM] -> 0 ] ;

```

103 States
2368 Transitions



Morphology



Declensions
Conjugations



parl

e

stem	
cat	verb
conjugation	1
lex	parler

suffix	
cat	verb
conjugation	1
tense	pres
person	1 3
number	sing

word	
cat	verb
lex	?Lex
tense	?Tense
person	?Person
number	?Number



stem	
cat	verb
conjugation	?Conj
lex	?Lex

suffix	
cat	verb
conjugation	?Conj
tense	?Tense
person	?Person
number	?Number

parle

word	
cat	verb
lex	parler
tense	pres
person	1 3
number	sing

One rule for all
finite verbs



Syntax

Agreement

Head features

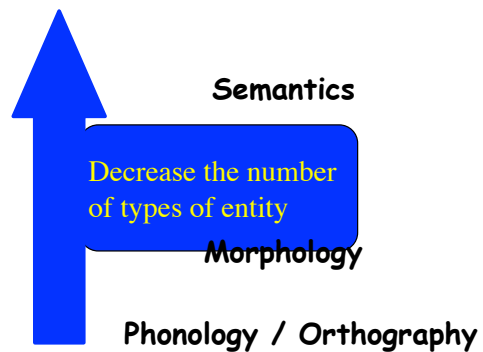
Subcategorization

Modification

Long-distance dependencies



Language Abstraction



Language Abstraction

Semantics

Syntax

Morphology

tying = tie+ing loved = love + ed

Language Abstraction

Semantics

Syntax

*loves = love+pl, love+3sg
children = child+pl*

Phonology / Orthography

Language Abstraction

Semantics

*The dog chased the cat = the
cat was chased by the dog*

Morphology

Phonology / Orthography

Language Abstraction

*Brutus killed Caesar =>
Caesar is dead*

Syntax

Morphology

Phonology / Orthography

Language Abstraction



Semantics

Syntax

Morphology

Phonology / Orthography

masc, fem, neut
sing, plur
nom, acc, gen, dat
past, present, future
perfective,
imperfective
...

Orthography

Language Abstraction



Semantics

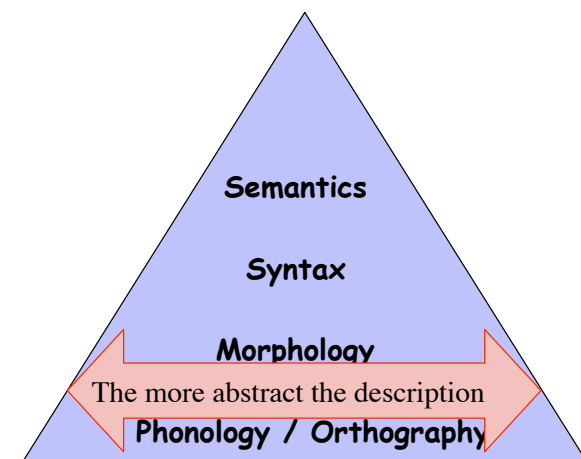
Syntax

Morphology

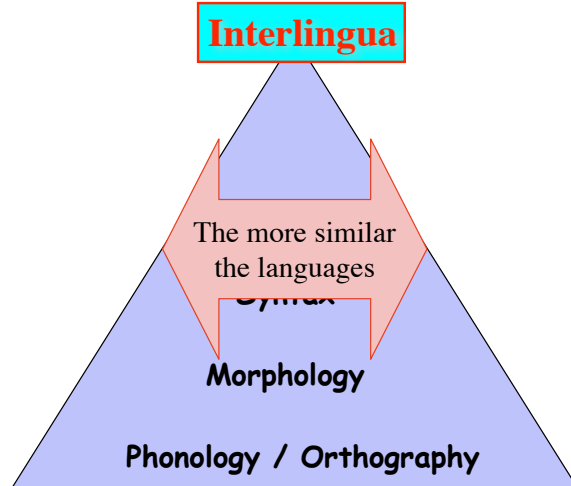
Phonology / Orthography

active, passive
decl, inter, imper
main, subordinate
subj, obj, mod
pos, neg
...

Vauquois Triangle



Vauquois Triangle

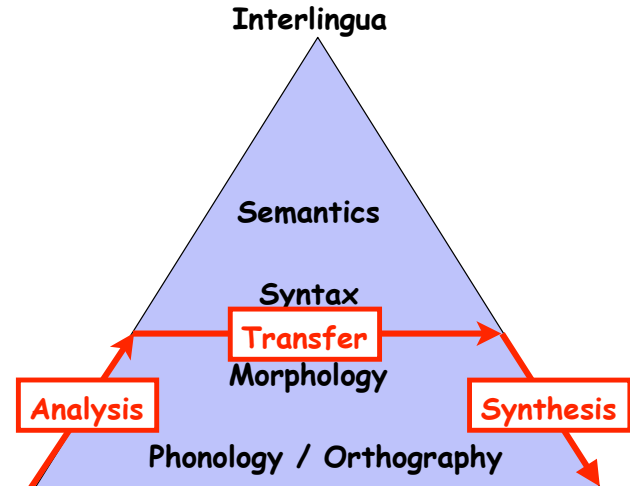


Martin Kay

Machine Translation

61

Vauquois Triangle

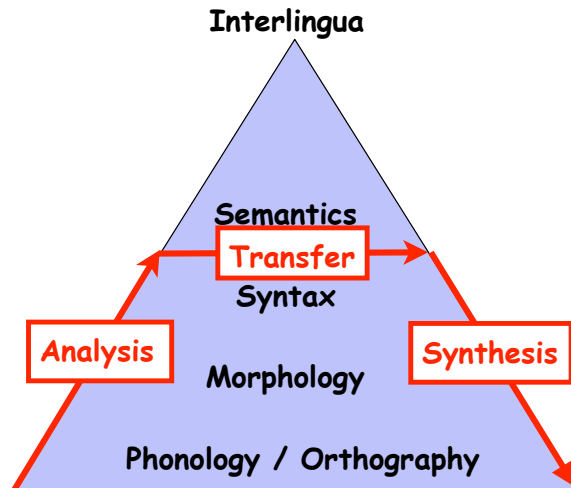


Martin Kay

Machine Translation

62

Vauquois Triangle

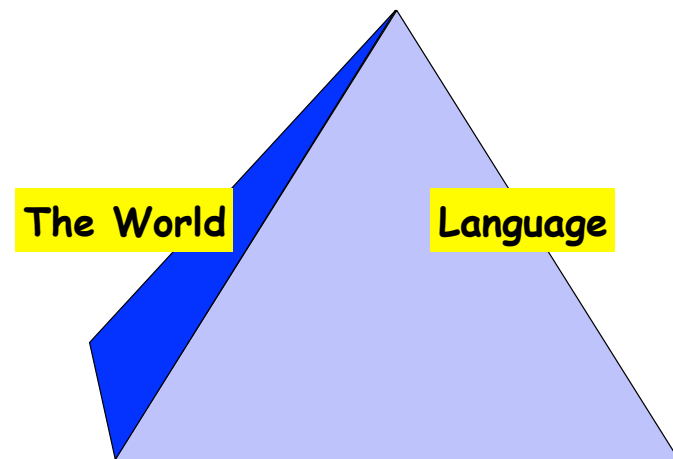


Martin Kay

Machine Translation

63

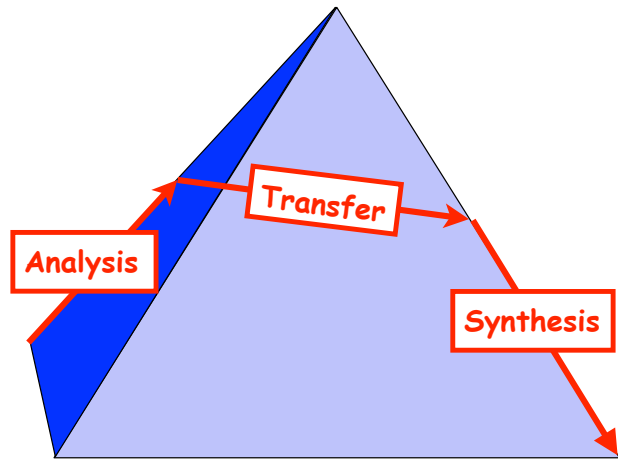
L'Arbitraire du Signe



Martin Kay

Machine Translation

64



Martin Kay

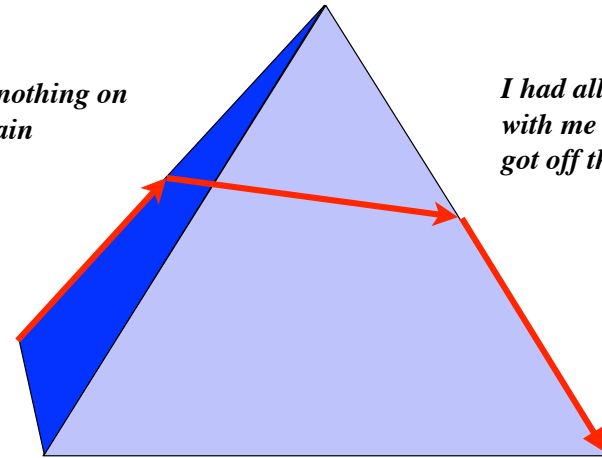
Machine Translation

65

L'Arbitraire du Signe

*I left nothing on
the train*

*I had all my stuff
with me when I
got off the train*



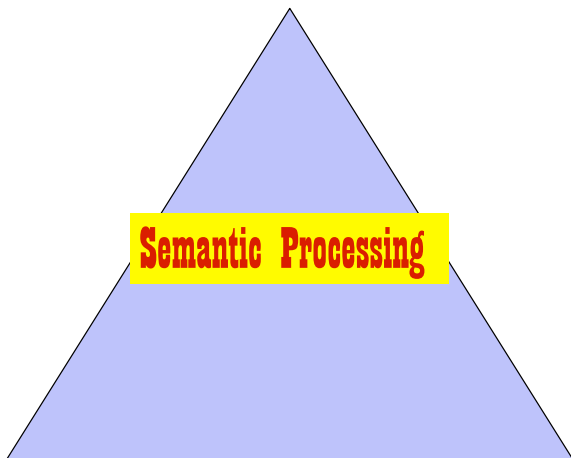
Martin Kay

Machine Translation

66

Vauquois Triangle

Semantic Processing



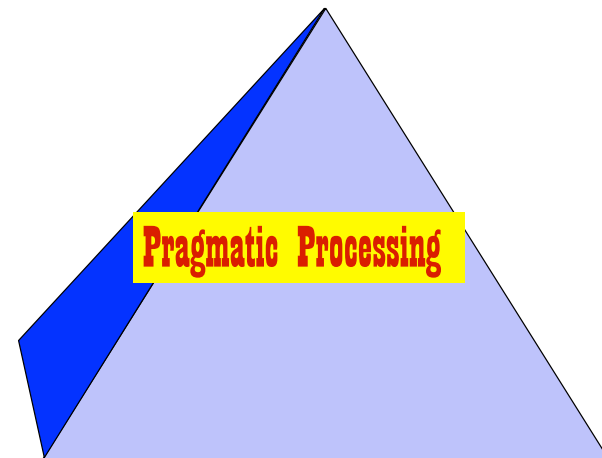
Martin Kay

Machine Translation

67

L'Arbitraire du Signe

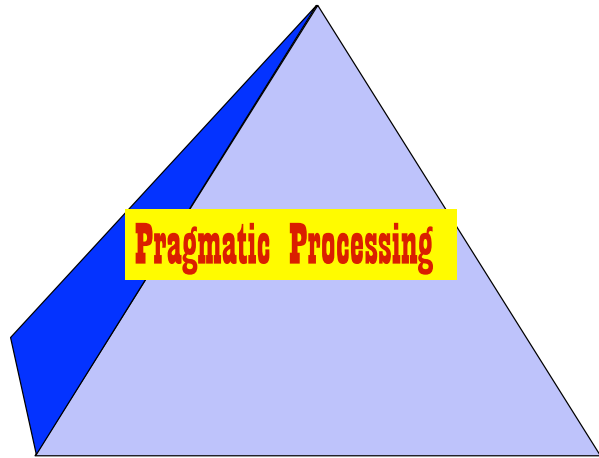
Pragmatic Processing



Martin Kay

Machine Translation

68



People, see everything

—in a context **Situated**

—from a point of view **Embodied**