Semantics
My Take

- Translation (Syntax and semantics)
- Morphology (Hebrew and Arabic)
- Shallow semantics (Shallow and deep)
TOW-AWAY
NO STOPPING
ANY TIME

REMOLQUE LEJOS
NO PARAR
CUALQUIER HORA
わたしな「おばあさん」です。あなたは？
watashi wa
"Obaasan" desu.
Anata wa?
Stats

- In 1952, no more than $10K was spent on MT. (Bar Hillel)
- In the US, $1.5M was spent in 1958 upon research connected with MT. (Bar Hillel)
- About $10B is spent annually on human translation. (Kevin Knight)
- About $50B is spent annually on human translation. (Frank Rudzicz)
The translator should first try to grasp the sense of the subject thoroughly, and then state the theme with perfect clearness in the other language. This, however, can not be done without changing the order of the words, putting many words for one word, or vice versâ, and adding or taking away words, so that the subject be perfectly intelligible in the language into which he translates.
I made her duck.
Bar Hillel’s Example

Little John was looking for his toy box.
Finally, he found it.
The box was in the pen.
John was very happy.
Bar Hillel (1960)

Fully automatic, high quality translation is not a reasonable goal, not even for scientific texts. A human translator ... is often obliged to make intelligent use of extra-linguistic knowledge which sometimes has to be of considerable breadth and depth. Without this knowledge he would often be in no position to resolve semantical ambiguities.
At present no way of constructing machines with such a knowledge is known, nor of writing programs which will ensure intelligent use of this knowledge.

Reasonable goals are then either fully automatic, low quality translation or partly automatic, high quality translation. Both are theoretically feasible....
Mettant en son dictionnaire un seul chiffre qui se rapporte à *aymer, amare, philein, et tous les synonymes* le livre qui sera écrit avec ces caractères pourra être interprété par tous ceux qui auront ce dictionnaire.
History

- 1950s: Intensive research
- 1960s: Direct replacement
- 1966: Funding cut
- 1975-1999: Resurgence
- 2000-2010: Statistical supremacy
- 2011-: ???
Spoken words are the symbols of mental experience and written words are the symbols of spoken words. Just as all men have not the same writing, so all men have not the same speech sounds, but the mental experiences, which these directly symbolize, are the same for all, as also are those things of which our experiences are the images.
Semantics

- Word level
- Sentence level
- Discourse level
Interlingua

- Semantic network (Richens, 1956; Collins+, 1969)
- Conceptual dependency; MOPTRANS (Schank, 1972)
- UNITRAN (Dorr, 1990)
- KBMT (Nirenburg&Carbonell+, 1992)
- Text Meaning Representation (Mahesh&Nirenburg, 1995)
- Lexical Conceptual Structure (Habash+, 2003)
שתיתי ولا אכלתי.
Examples

- She picked dates.
- She picked plump dates from the palm.
She picked plump dates from the palm.

 nieuwe haarijken mammena

She lifted [calendar] dates plump [woman] from palm [of hand].
The cute woman lost her beautiful old green leather bag.
The cute woman lost her beautiful old green leather bag.

האישה חמודה איבדה התיק יפה וющהך

הישחק שלה עור יוקר
The cute woman lost her beautiful old green leather bag.

The woman lost cute beautiful the bag old her leather green.
The cute woman lost her beautiful old green leather bag.

AFRIKAANS: Die oulike vrou verloor haar pragtige ou groen leer tas.
ALBANIAN: Gruaja e bukur humbi çantën e bukur e saj të vjetër e gjelbër lëkurë.
ARABIC: فقدت زوجة جميلة لها حقيقة جلدية خضراء جميلة القديمة.
BELARUSIAN: Я страціў прыгожая жонка прыгожы зялёны скураны мяшок старых.
BULGARIAN: Загубих си красива жена, красива зелена кожена чанта от стари.
CATALAN: Vaig perdre la meva bella dona, bella borsa de cuir verd d'abans.
CHINESE......: 我失去了我美丽的妻子，美丽的绿色皮包前.
CROATIAN: Izgubio sam lijepu ženu, prekrasan prednji džep.
ENGLISH: I lost my beautiful wife, beautiful front pocket.
The cute woman lost her beautiful old green leather bag.

Wakanari green leather bag, the old woman had lost her beauty.
Jean François Champollion
Rosetta Stone
Coptic
Egyptian Morphology

- Roots (2-3 consonants)
- Verbs: person, aspect, mood, voice
- Enclitic pronouns
- Glyphs represent 1-3 consonants
Given life, stability, and dominion – Joy of Ra, as Ra forever.
English Morphology

- eat, eats, ate, eaten, eating, eater, eatable, eatables, eat, eats

- 315,000 lemmata

- 988,968 (est.) words
Morphology

- French verbs: 3 persons; 2 numbers; 2 genders; 2 voices; 10 tenses; 6 modes

- German: haltend, halte, halten, haltet, haltest, haeelt, haeeltst, hielte, hielten, hieltet, hieltest, ..., hintanzuhhalten, ..., zusammenhalten, ....

- Russian: delayet, peredelayet, delayu, delayesh, delayem, delayut, delal, delali, delala, ..., sdelayet, dodelayet, dodelal, sdelal, sdelayu, dodelayu, ....
Arabic Morphology

- 5-10,000 roots
- 114,000 lemmata
- 3,000,000 forms (inflected, w/o clitics)
- 500,000,000 words
- 60,000,000,000 words (!)
- Average 6.8 meanings per word
DID YOU KNOW THAT THE ENGLISH HAVE OVER 400 WORDS FOR NO?

ENGLISH HAVE 10 TIMES AS MANY NEGATIVE WORDS.
Do you believe...

There are 160 words in Arabic for camel.
Buckwalter’s Lemmatizer

- 40K lemmata
- Try all segmentations into prefix-stem-suffix
- Lists of legal prefixes, stems, suffixes
- Manually inflected
- English glossary
Hebrew Lexicon (1975)

- Noun: Gloss; Root; Gender/Number; Articles; Prepositional proclitics; Pronominal enclitics
- Verb: Tenses; Persons; etc.
- Based on Even Shoshan
- 3 linguist-years
Hebrew Morphology

- 3-5,000 roots
- 35,000 lemmata (incl. 10% international)
- 2,500,000 naked words (inflected+suffixes)
- 100,000,000 words (w/ proclitics)
- 20,000 words for one verb
- Average >4 meanings (homographs) per word
Hebrew Vocabulary

The Hebrew vocabulary tree shows that Hebrew morphology is characterized by a small number of d-forms and a large number of derived kernels, compounds, and l-forms. About 35,000 kernels, about 500,000 ob do, and 209,000/ob about 500,000 derived kernels, compounds, and l-forms are present.

The numbers show that Hebrew morphology is characterized by a small number of d-forms and a large number of derived kernels, compounds, and l-forms.

Some of the derivations of the root RAH (to see):

- RAH (saw)
- MRAH (mirror)
- RAYWN (interview)
- HTRAH (he met)
- RAYTY (I saw)
- YRAW (they will see)
- RAYT (you saw)
- MRAWT (mirrors)
- RAYWNWT (interviews)
- TTRAW (you will meet)
- RAYTYK (I saw you)
- YRAWNY (they will see me)
- MRAWTYKN (your mirrors)
- TTRAW (you will meet)
- KŠMRAWTYKN (and when your mirrors)
- KŠTTRAW (when you will meet)
- WTTRAW (and you will meet)

It should be pointed out that most of the forms are highly homographic. Only one of the possible meanings is given in the translations on the tree, owing to the adjunction of prefixes, suffixes, and infixes and the deletion of other letters during inflection.

The inflected nature of the language and the omission of all vowels in most of the Hebrew texts (in particular, in all of the Responsa literature) induce a very large number of homographs: about four per form on the average (and in extreme cases up to a few tens).

5. Synthesis, Analysts, and Their Fusion

As indicated in Section 2, both analysis and synthesis may be used to create the grammatical relations of the vocabulary tree.

A feasibility study of morphological analysis of Hebrew (see [4, 8]) indicated that such a process is feasible, but requires very accurate manually prepared dictionaries and quite delicate computer programs that would be too cumbersome and lengthy for practical implementation.

The (dynamic) synthesis approach, which is easier to implement and less sensitive, has therefore been preferred. Algorithms were formulated for the automatic generation of the set $F(s)$ of all 1-forms derivable from any given d-form $s$, requiring the user to...
Nominal Morphology

dughter bt

20 derived forms (e.g. البنנות bj daughters–our)

>1000 forms (e.g. והנו ובנהנו and–when–daughters–our)
לבנה
## Lexical Ambiguity

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. brick</td>
<td>12. to a builder</td>
</tr>
<tr>
<td>2. her brick</td>
<td>13. to the builder</td>
</tr>
<tr>
<td>3. moon</td>
<td>14. to her builder</td>
</tr>
<tr>
<td>4. her moon</td>
<td>15. to a beaver</td>
</tr>
<tr>
<td>5. kefir</td>
<td>16. to the beaver</td>
</tr>
<tr>
<td>6. frankinsense</td>
<td>17. to her beaver</td>
</tr>
<tr>
<td>7. her frankinsense</td>
<td>18. to wisdom</td>
</tr>
<tr>
<td>8. her bleach</td>
<td>19. to the wisdom</td>
</tr>
<tr>
<td>9. she bleached</td>
<td>20. to her wisdom</td>
</tr>
<tr>
<td>10. was bleached</td>
<td></td>
</tr>
<tr>
<td>11. to her son</td>
<td>21. Livne ...</td>
</tr>
</tbody>
</table>
"Oh! Four steps to the left and then three to the right! ... What kind of a dance was I doing?"
Parallel Corpora
Saul questionnaires answers that ethical arguments against the claims of that share of continuing our religion, philosophy and his followers for other religions and from species of Israel.

I was asked to state what arguments and replies I could bring to bear against the attacks of philosophers and followers of other religions, and also against sectarians who attacked the rest of Israel.
(1) \( S \rightarrow \text{verb} \cdot O \cdot C \iff S' \rightarrow \text{verb}', O', C' \in W \),
\( S, S' \in W_X, O, O' \in W_Y, C, C' \in W_Z \)
where \( W_X, W_Y, \) and \( W_Z \) are semantic groups of words \( X, Y, Z \).

(2) A man eats vegetables.

(3) 人は 野菜を たべる。
\((\text{man} \ (\text{vegetable} \ (\text{eat}))\)

(4) He eats potatoes.

(5)  man 他
    vegetable  もも

(6) 人 他
    (man)  (he)
    野菜 じゃがいも
    (vegetable) (potato)

(7) 他は じゃがいもを たべる。
WordNet

Visuwords v2.12
Experiment: Arabic

- Example-based translation
- Extract potential nominal synonyms (based on Buckwalter's glossary)
- Use in matching step
- More (verbs) before lunch
We believe that competent translation of a connected text is impossible unless the text has first been understood, in some reasonably deep sense.
Semantic Roles

- **Agent** (Alice gave the key to Bob.)
- **Source** (Alice gave the key to Bob.)
- **Theme** (Alice gave the key to Bob.)
- **Recipient** (Alice gave the key to Bob.)
- **Patient** (Alice locked the door for Bob with her key.)
- **Benefactee** (Alice locked the door for Bob with her key.)
- **Instrument** (Alice locked the door for Bob with her key.)
- **Force** (The wind made Alice cold.)
- **Experiencer** (The wind made Alice cold.)
- ...
Semantic Resources

- VerbNet
- FrameNet
- Logical Conceptual Structures (LCS)
- PropBank has some annotations
- OntoNotes

...
Semantic Parsing

- Parse and assign roles
- Senseval 3: Determine semantic roles
Cases

- Nominative: subject
- Accusative: direct object
- Dative: indirect object

Some languages (e.g. Greek, Latin, Russian) indicate roles via morphology
VerbNet

- Large hierarchical lexicon for English verbs
- Linked to WordNet
- Organized into verb classes
- Each class
  - thematic roles
  - selectional restrictions on arguments
  - syntactic description
Hit

Roles and Restrictions:

- Agent [+int_control]
- Patient [+concrete]
- Instrument [+concrete]

Members: bang, bash, hit, kick, ...
Hit

Frame: (Basic Transitive)

Example: Paula hit the ball

Syntax: Agent V Patient

Semantic:

- cause(Agent, E)
- manner(during(E), directed movimiento, Agent)
- !contact(during(E), Agent, Patient)
- manner(end(E), forceful, Agent)
- contact(end(E), Agent, Patient)
Restrictions

- concrete
  - solid
  - shape
  - time
  - state
  - abstract
  - scalar
  - currency
  - location
  - organization
- int-control
  - force
  - animate
  - natural
  - plant
  - phys-obj
  - comestible
  - artifact
  - rigid
  - pointed
  - non-rigid
  - elongated
- SelRestr
  - idea
  - sound
  - communication
  - regionPP
  - place
  - object
  - tool
  - garment
  - machine
  - body-part
  - animal
  - human
  - vehicle
Linkage

- Restrictions in VN can be mapped to WN

- Not simple
  - I sipped my tea.
  - My cat likes to lap milk.
Transformer, Ltd.

- English to Hebrew
- High quality, general purpose MT
- Semantic parsing of English
- Comprehensive ontology of verbs and their theta roles with restrictions to refined WN categories
- Slow

Tal Kedar; Yaacov Zoaretz ...
Theta Roles

- Agent (Alice gave the key to Bob.)
- Theme (Alice gave the key to Bob.)
- Recipient (Alice gave the key to Bob.)
- ...

Agent (Alice gave the key to Bob.)
Theme (Alice gave the key to Bob.)
Recipient (Alice gave the key to Bob.)
...
What can I eat?

- I ate hamburgers at the city center yesterday.
- ...  
- I eat hamburgers at the center of the city yesterday.
- 我吃了汉堡市中心的昨天。
- I ate the center of Hamburg yesterday.
Edible Fruit
Picking in Hebrew

- Olives
- Grapes
- Figs, myrrh, spices
- Dates
- Other fruit

- масק
- בצר
- ארה; עזר
- גזר
- קטף
Machine Learning

- Learn selectional restrictions
- Use large corpora
Garden Path

The man who hunts ducks out on weekends.
Greedy Parsing

- Bradley Pritchett (1992)
- David Schneider (1999)
- Uriel Cohen Priva (2006)
As for trans. 4 Essais, again I’m not anxious. I’m not anxious. I'm not fast about these things, and it wd. take me hours to translate a phrase like 'des formes', or 'ces formes', to satisfy my conscience and to feel that I’m doing more than justice to René.
Hybrids

THE BIONIC MAN AT A CUT PRICE

- BIONIC EYE used to bring back limited vision to those who have lost their sight. With an implant on the retina infra-red sight (below) could be possible.
  Cost: £18,000

- BIONIC STRENGTH (Berkeley Blonics exoskeleton).
  Cost: £12,000

- BIONIC LEGS (Ossur power knees). Works as an extension of the user and synchronises with good leg.
  Cost: £76,000

- BIONIC HAND OR I-LIMB (above) is attached to nerves in the arm to pick up signals from the brain. It can be covered with artificial skin.
  Cost: £44,000

Source: E&T magazine

Lee Majors as television's Six Million Dollar Man.
Thank You