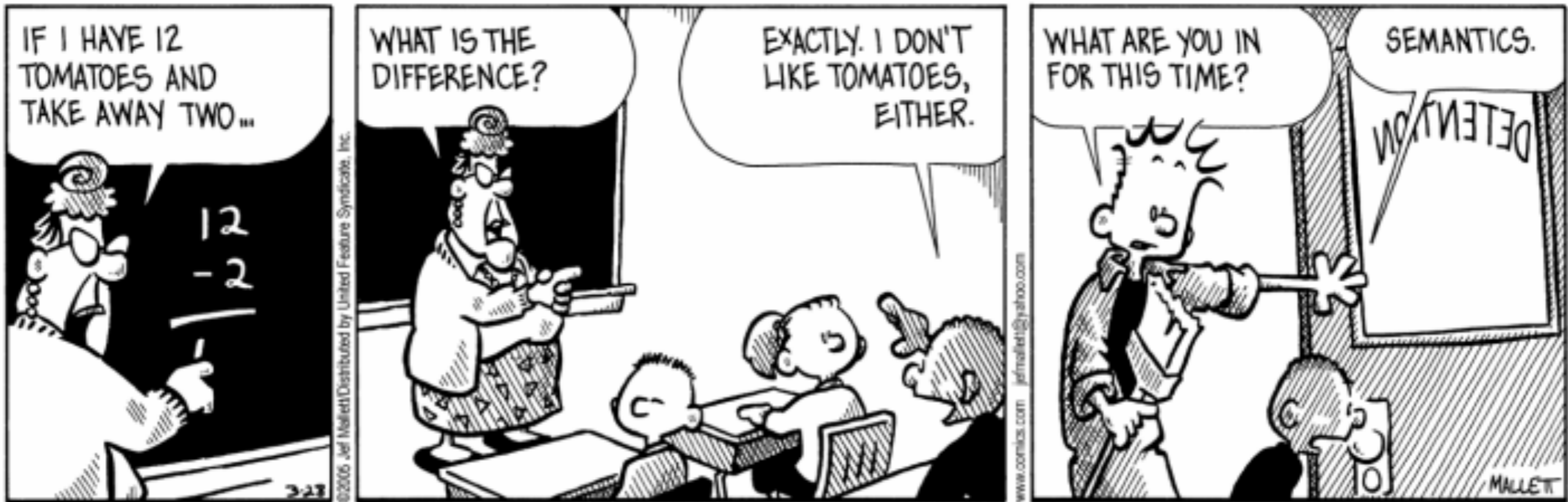


Semantics





My Take

- Translation (Syntax and semantics)
- Morphology (Hebrew and Arabic)
- Shallow semantics (Shallow and deep)



THORCHTE - WAS
DOEN - NIET - VALLEN

• BABELON •
• GENESIS - 11 •



English to Spanish

TOW-AWAY
NO STOPPING
 ANY
 TIME

FOR TOWED CARS
 PHONE 553-1235



English to Spanish

REMOLQUE LEJOS
NO PARAR
 CUALQUIER
 HORA

FOR TOWED CARS
 PHONE 553-1235



わたしは「おばあさん」
です。 あなたは？

watashi wa
"Obaasan" desu.
Anata wa?

אני אובָּסָן,
ואתה?



Stats

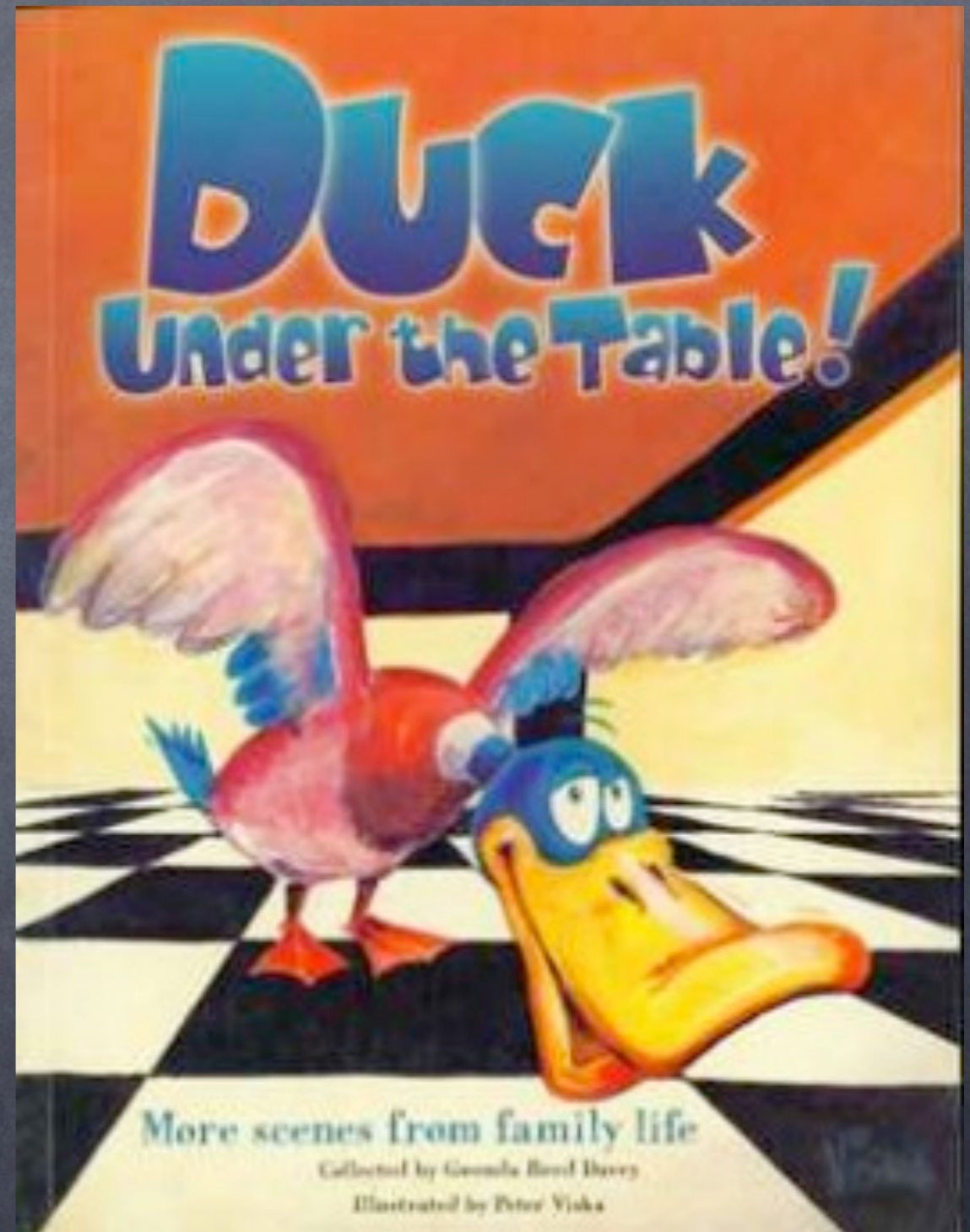
- In 1952, no more than \$10K was spent on MT. (Bar Hillel)
- In the US, \$1.5M was spent in 1958 upon research connected with MT. (Bar Hillel)
- About \$10B is spent annually on human translation. (Kevin Knight)
- About \$50B is spent annually on human translation. (Frank Rudzicz)

Maimonides (1199)

- 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

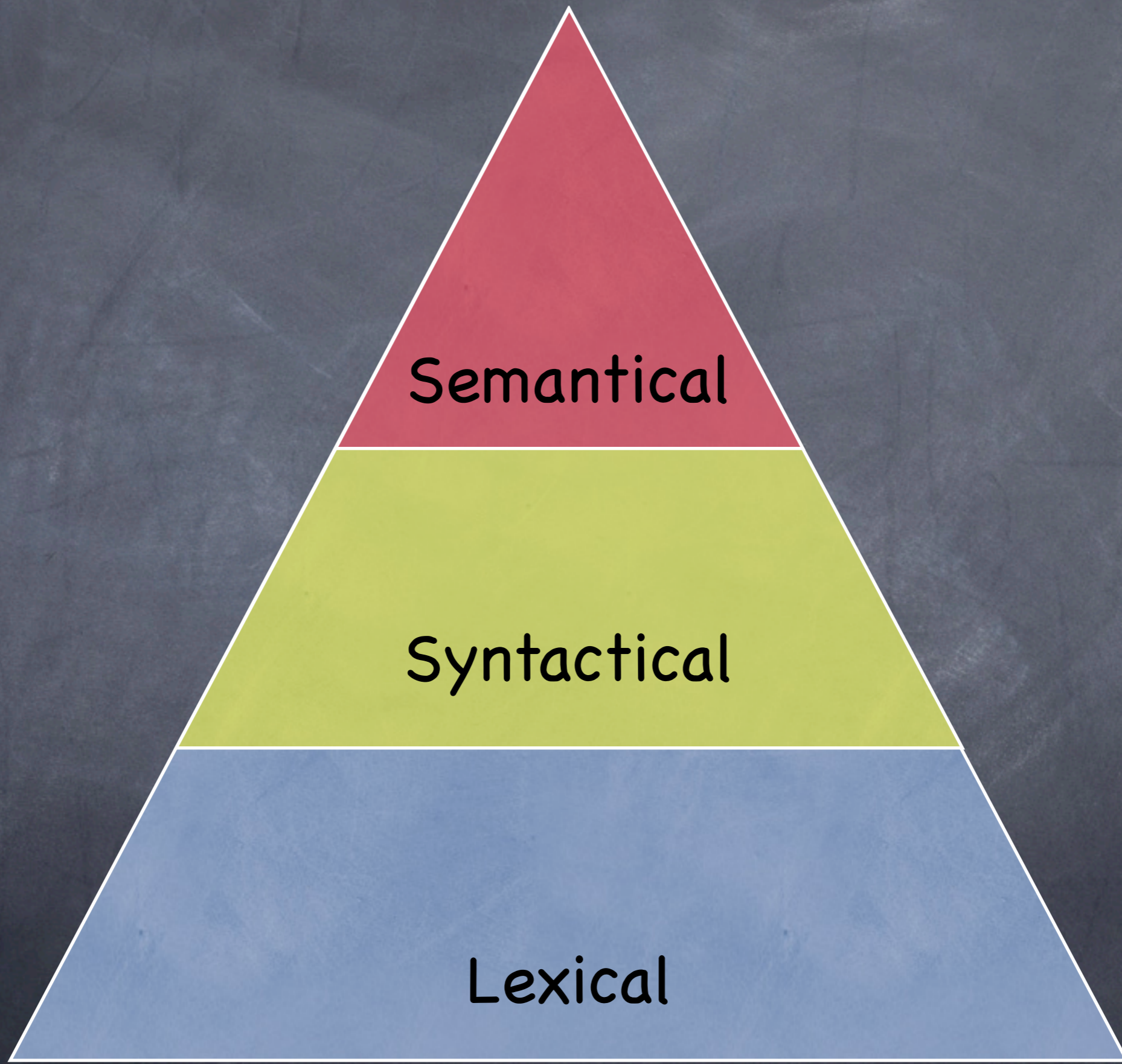
- Little John was looking for his toy box.
- Finally, he found it.
- The box was in the pen.
- John was very happy.

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.

Bar Hillel (1960)

- 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...



Semantical

Syntactical

Lexical

Descartes (1629)

- 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–: ???

Aristotle (-350)

- 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)



Translate

אכלתי ולא שתיתי

Hebrew to English translation

I ate and drank

[Listen](#)

Try a new browser

Download

Examples

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

Googlese

From: English ▼

To: Hebrew ▼

Translate

English to Hebrew translation

She picked plump dates from the palm.

היא הרימה תאריכים שמנמנה
מכף.

She lifted
[calendar] dates
plump [woman]
from palm [of hand].

Googlese

The cute woman lost her beautiful old green leather bag.

Googlese

The cute woman lost her beautiful old green leather bag.

האישה איבדה חמוד יפה התיק
הישן שלה עור ירוק.

Googlese

The cute woman lost her beautiful old green leather bag.

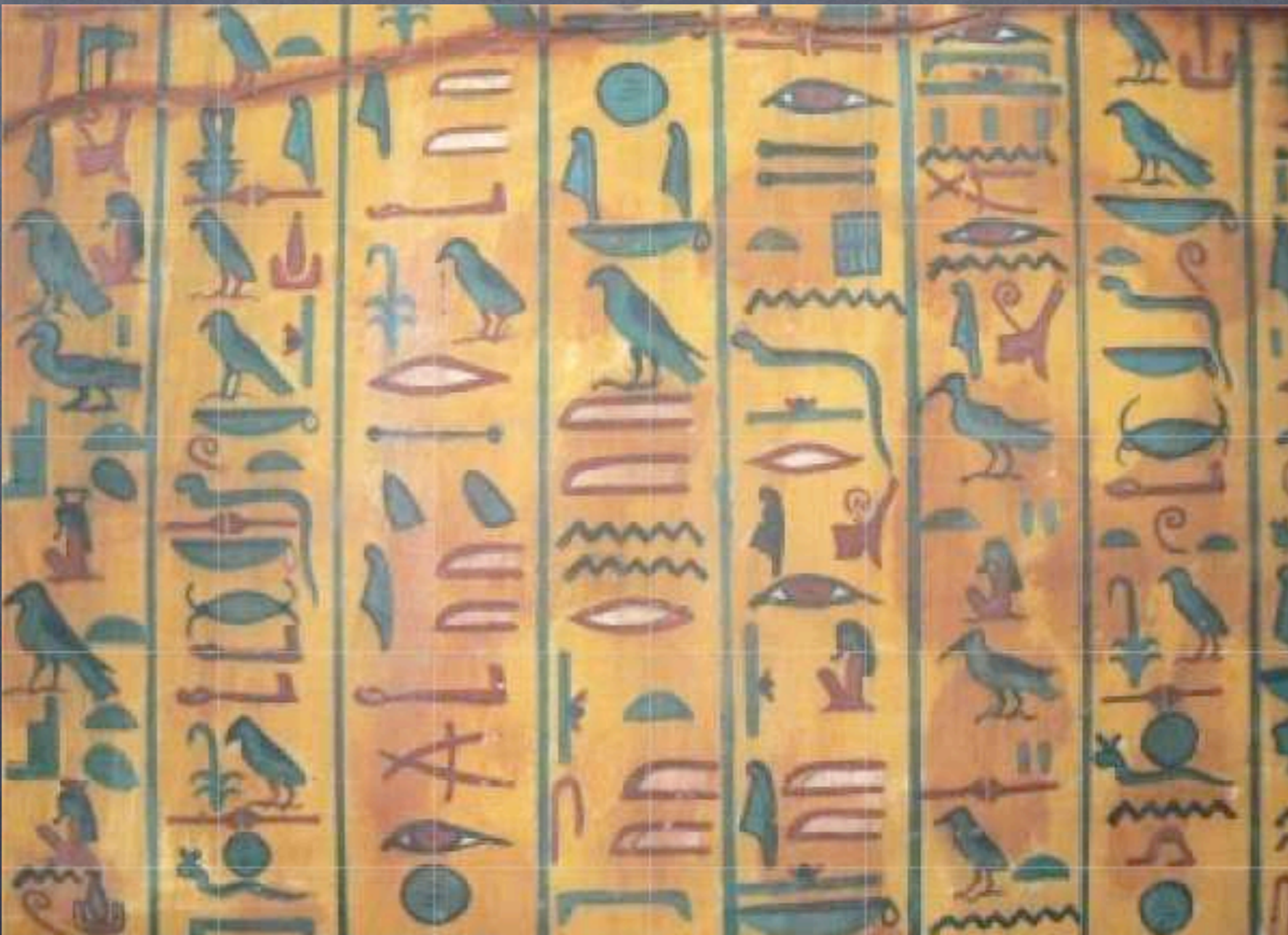
The woman lost cute beautiful the bag old her leather green.

Bad Translator Site

- 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 حقيبة 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.

Translation Party Site

- The cute woman lost her beautiful old green leather bag.
- Wakanari green leather bag, the old woman had lost her beauty.



Jean François Champollion





Coptic

Egyptian Morphology

- Roots (2–3 consonants)
- Verbs: person, aspect, mood, voice
- Enclitic pronouns
- Glyphs represent 1–3 consonants

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, haelt, haeltst, hieltst, hielte, hielten, hieltet, hieltetst, ... , hintanzuhalten, ... , 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 ?



Search ID: gth02229

© Original Artist

Reproduction rights obtainable from

www.CartoonStock.com

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

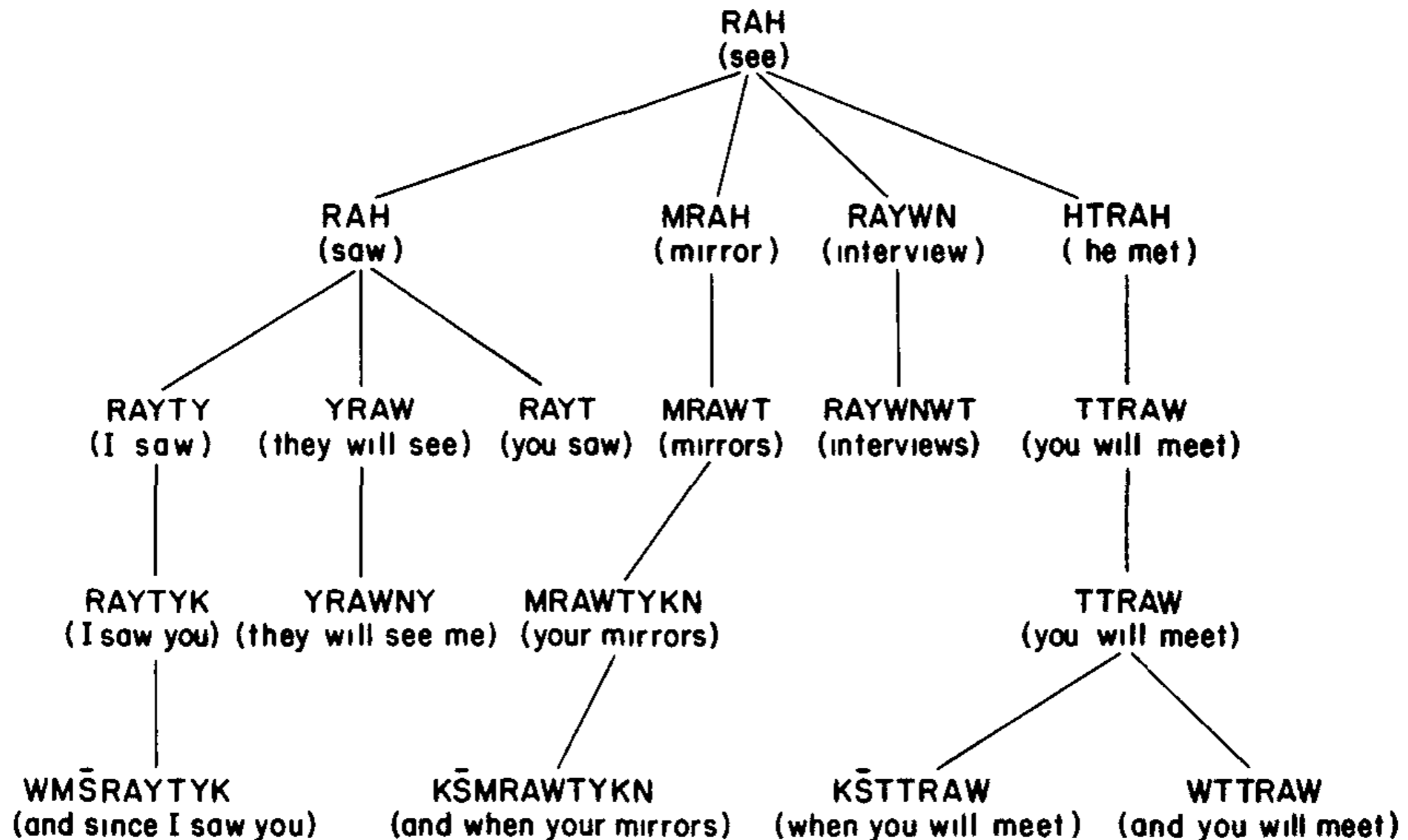
Hebrew Lexicon (1975)

- 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 Morphology



Nominal Morphology

- daughter בת
- 20 derived forms (e.g. בנותינו daughters-our)
- >1000 forms (e.g. וכשבנותינו and-when-daughters-our)

לבנה



Lexical Ambiguity

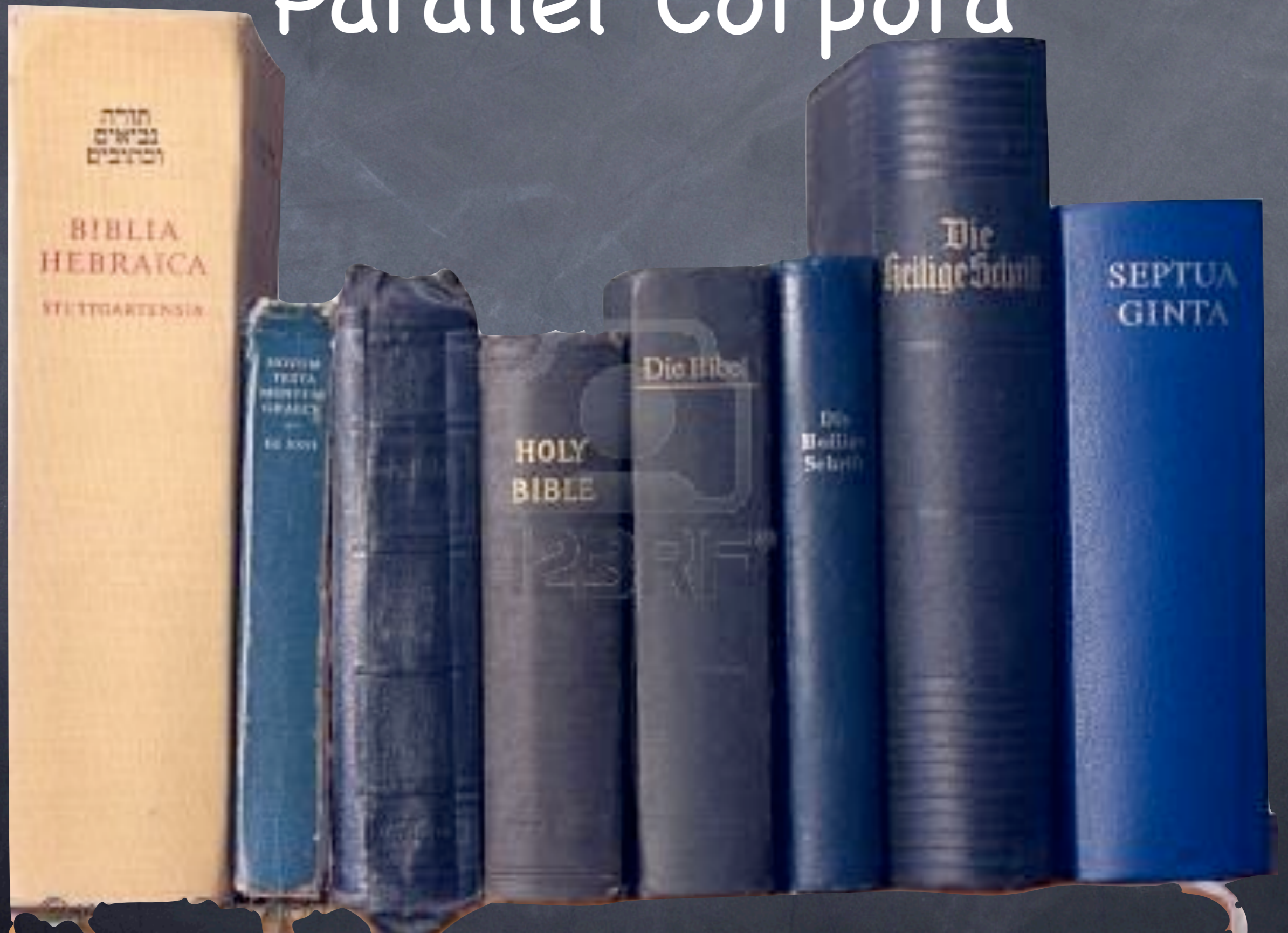
1. brick
2. her brick
3. moon
4. her moon
5. kefir
6. frankinsense
7. her frankinsense
8. her bleach
9. she bleached
10. was bleached
11. to her son
12. to a builder
13. to the builder
14. to her builder
15. to a beaver
16. to the beaver
17. to her beaver
18. to wisdom
19. to the wisdom
20. to her wisdom
21. Livne ...

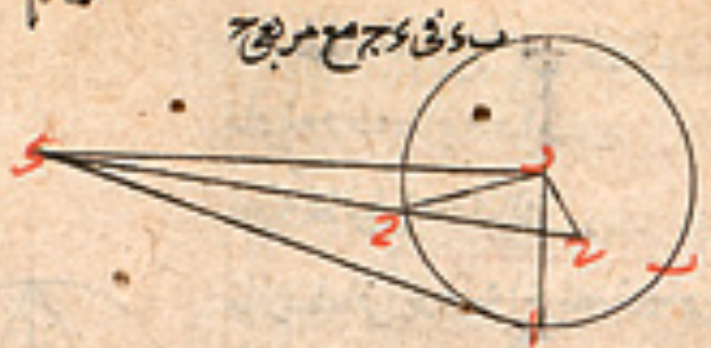
THE FAR SIDE : SYNTAX VS. SEMANTICS



"Oh! Four steps to the left and then three to the right! ... What kind of a dance was I doing?"

Parallel Corpora



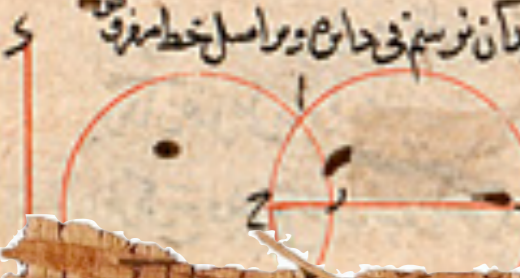


ح وشتد كما صار سطح
 ح ح راعني مربع رح بل
 مربع راسا ويا المربعي ح و
 ح راعني مربع ره ولكن سطح

ب د في ح مساوي مربع و ا ف ب ا و ا و ليا و ان مربع ر د ف ا و ب ا و
 قائمه و د ا ق ماس و اختلاف الوقوع على قياس الشكل المتقدم المقالة
 الرابعة ستة عشر شكلا صد ر ا ذ الحاط شكل بسكني ح س ن م ا س

المقالة الرابعة

ن د ا ب ا الحاط اضلاع الحاط اسند الحاط الى الحاط بانه فيه والمحط
 الى الحاط بانه يملكه الا لشكال بر يدان ن ر س م في دائر و بر اسل خط ممزق



لكن اطول من قطرهما مثلا في
 دائر ا ب ح مثل خطك فيخرج

فها قطر وهو ح و
 ومثل د ه

Fragment of a manuscript page with Arabic script, partially obscured by a tear. The text is written in a cursive hand and appears to be a continuation of the geometric treatise.



Præclarissimus liber elementorum Euclidis periphrasice
 explicatum: in artem Geometricæ incipit quæsoelicolline:



Unctus est cuius po no est. **L**inea est
 longitudo sine latitudine cuius quidē ex-
 tremitates si duo pōcta. **L**inea recta
 ē ab vno pōcto ad aliū brevissima extē-
 sio i extremates suas vtrūq; eorū reci-
 piens. **S**upficies ē q̄ longitudo & lati-
 tudine tria bycorū termi quidē sūt linee.
Supficies plana ē ab vna linea ad a-
 liā extēsiō i extremates suas recipiēs
Angulus planus ē duarū linearū al-
 ternis pōctis: quaz extēsiō ē sup sup-
 ficie applicatioq; nō directā. **Q**uādo aut angulum pōtinet due
 linee recte rectiline⁹ angulus noiat. **A**ngulus
 si recta linea sup rectā
 fuerit duosq; anguli vtrobiq; fuerit eque-
 cos: eorū vterq; rect⁹ erit
Lineaq; linee supflās ei cui supflāt p-
 pendicularis vocat. **A**ngulus vō qui
 recto maior ē obtusus dicit. **A**ngulus
 vō minor recto acut⁹ appellat. **T**ermin⁹
 ē qd vnicuiusq; termin⁹ ē. **F**igura
 ē q̄ termin⁹ vltimis pōtinet. **C**ircul⁹
 ē figura plana vna quōm li-
 nea pōcta: q̄ circūferentiā noiat: in cui⁹
 medio pōct⁹ ē: a quo oēs
 linee recte ad circūferentiā extētes
 libitimes sūt equales. **E**t hic
 quidē pōct⁹ cētū circuli dē. **D**iameter
 circuli ē linea recta que
 sup ei⁹ cētū: trāsfiēs extrematesq; suas
 circūferentiē applicans
 circulū i duo media diuidit. **S**emirculus
 ē figura plana dia-
 metro circuli & medietate circūferentiē
 pōcta. **P**ortio circuli
 ē figura plana recta linea & parte
 circūferentiē pōcta: semicirculo
 quidē aut maior aut minor.
Rectilinee figure sūt q̄ rectis li-
 neis cōtinentur. **T**riangul⁹
 triangul⁹ ē figura plana tri-
 latera q̄ trib⁹ rectis lineis: quedā
 lineis. q̄dā miltalere que pluribus
 lineis. **F**igurarū tri-
 laterarū: alia
 equalia. **A**lia triangulos duo
 būs
 in equalium laterū. **I**say iterū
 unū angulum habens. **A**lia ē an-
 gulum habens. **A**lia est origoni-
 ti. **F**igurarū autē quadrilateraz
 aliterū atq; rectangulū. **A**lia est
 ctangula: sed equilatera non est.
 quilatera: sed rectangula non est.

De principijs p se notia: e pno de diffini-
 tionibus eandem.



Judeo-Arabic

סילת עמא ענדי מן אלאחתגאג עלי מכ'אלפינא מן אלפלאספה ואהל
אלאדיאן ת'ם עלי אלכ'וארג אלד'ין יכ'אלפון אלגמהור

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.

שאול שאלוני לטענות ולתשובות אשר אתי נגד טענותיהם
של החולקים על דתנו מבין הנמשכים אחר הפילוסופיה
ומבין מאמיני דתות אחרות ומבין המינים מבני ישראל

Example Based (1984)

(1) $S \cdot v e r b \cdot O \cdot C \longleftrightarrow S' \text{ は} \cdot O' \text{ を} \cdot C' \text{ に} \cdot v e r b'$,

$S, S' \in W_X, \quad O, O' \in W_Y, \quad C, C' \in W_Z$

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。

(man) (vegetable) (eat)

(4) He eats potatoes.

(5) man ~ he

vegetable ~ potato

(6) 人 ~ 彼
(man) (he)

野菜 ~ ジャガイモ
(vegetable) (potato)

(7) 彼は ジャガイモを たべる。

Experiment: Arabic

- Example-based translation
- Extract potential nominal synonyms (based on Buckwalter's glossary)
- Use in matching step
- More (verbs) before lunch

Carbonell et al. (1981)

- 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

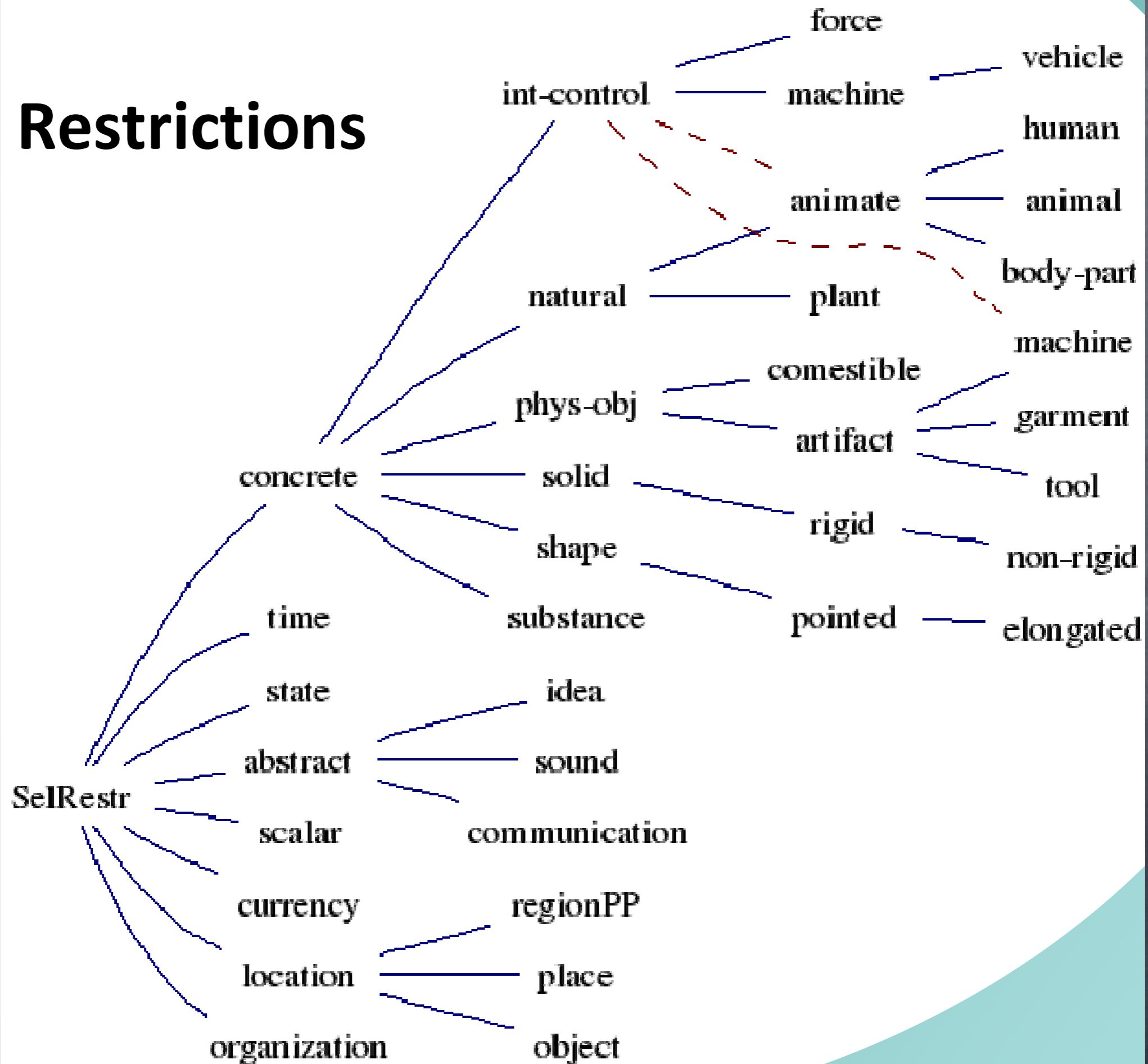
Hit

- 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:
 - $\text{cause}(\text{Agent}, E)$
 - $\text{manner}(\text{during}(E), \text{directedmotion}, \text{Agent})$
 - $\text{!contact}(\text{during}(E), \text{Agent}, \text{Patient})$
 - $\text{manner}(\text{end}(E), \text{forceful}, \text{Agent})$
 - $\text{contact}(\text{end}(E), \text{Agent}, \text{Patient})$

Restrictions



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**.)
- ...



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

👁 מסק

👁 בצר

👁 ארה; עדר

👁 גדר

👁 קטר

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)

Ezra Pound

- 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 Bionics exoskeleton).

Cost: **£12,000**

► **BIONIC LEGS** (Ossur power knees). Works as an extension of the user and synchronises with good leg.

Cost: **£76,000**



Lee Majors as television's Six Million Dollar Man.



► **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

**Thank
You**

Mahalo

Kiitos

Tack

Toda

Obbrigado

Grazie

Thanks

Takk

Gracias

Merci