

# Machine Translation

美国关岛国际机场及其办公室均接获一名自称沙地阿拉伯富商拉登等发出的电子邮件，威胁将会向机场等公众地方发动生化袭击後，关岛经保持高度戒备。



The U.S. island of Guam is maintaining a high state of alert after the Guam airport and its offices both received an e-mail from someone calling himself the Saudi Arabian Osama bin Laden and threatening a biological/chemical attack against public places such as the airport .

The classic acid test for natural language processing.

Requires capabilities in both interpretation and generation.

About \$10 billion spent annually on human translation.

Mainly slides from Kevin Knight (at ISI)

# Translation (human and machine)

对外经济贸易合作部今天提供的数据表明，今年至十一月中国实际利用外资四百六十九点五九亿美元，其中包括外商直接投资四百点零七亿美元。

**Ref** According to the data provided today by the Ministry of Foreign Trade and Economic Cooperation, as of November this year, China has actually utilized 46.959 billion US dollars of foreign capital, including 40.007 billion US dollars of direct investment from foreign businessmen.

**IBM4:**

the Ministry of Foreign Trade and Economic Cooperation, including foreign direct investment 40.007 billion US dollars today provide data include that year to November china actually using foreign 46.959 billion US dollars and

**Yamada/Knight:**

today's available data of the Ministry of Foreign Trade and Economic Cooperation shows that china's actual utilization of November this year will include 40.007 billion US dollars for the foreign direct investment among 46.959 billion US dollars in foreign capital

# Machine Translation History

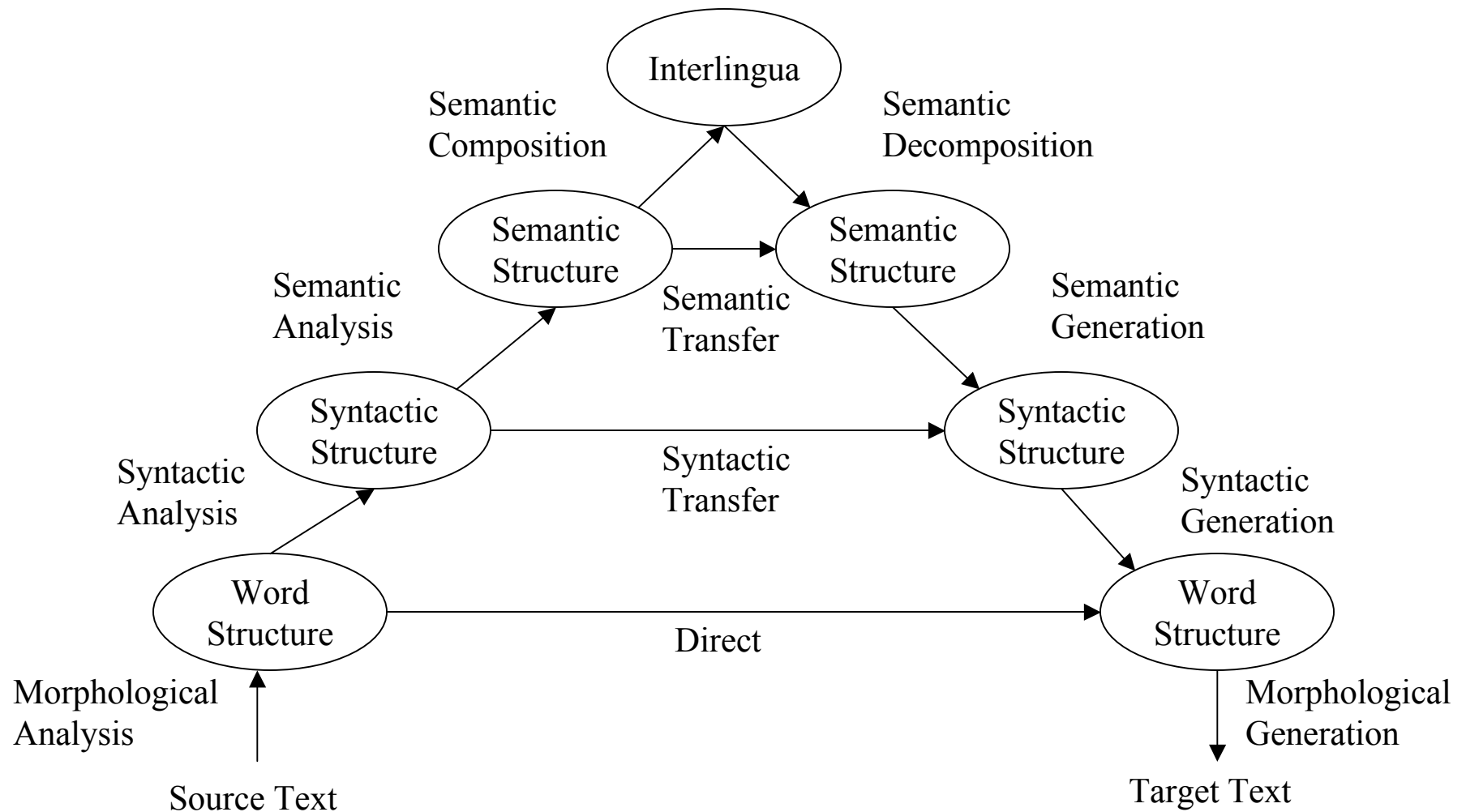
- 1950's: Intensive research activity in MT
- 1960's: Direct word-for-word replacement
- 1966 (ALPAC): NRC Report on MT
  - Conclusion: MT no longer worthy of serious scientific investigation.
- 1966-1975: 'Recovery period'
- 1975-1985: Resurgence (Europe, Japan)
  - Domain specific rule-based systems
- 1985-1995: Gradual Resurgence (US)
- 1995-2005: Statistical MT surges ahead

<http://ourworld.compuserve.com/homepages/WJHutchins/MTS-93.htm>

# What happened between ALPAC and Now?

- Need for MT and other NLP applications confirmed
- Change in expectations
- Computers have become faster, more powerful
- WWW
- Political state of the world
- Maturation of Linguistics
- Development of statistical and hybrid statistical/symbolic approaches

# Three MT Approaches: Direct, Transfer, Interlingual (Vauquois triangle)



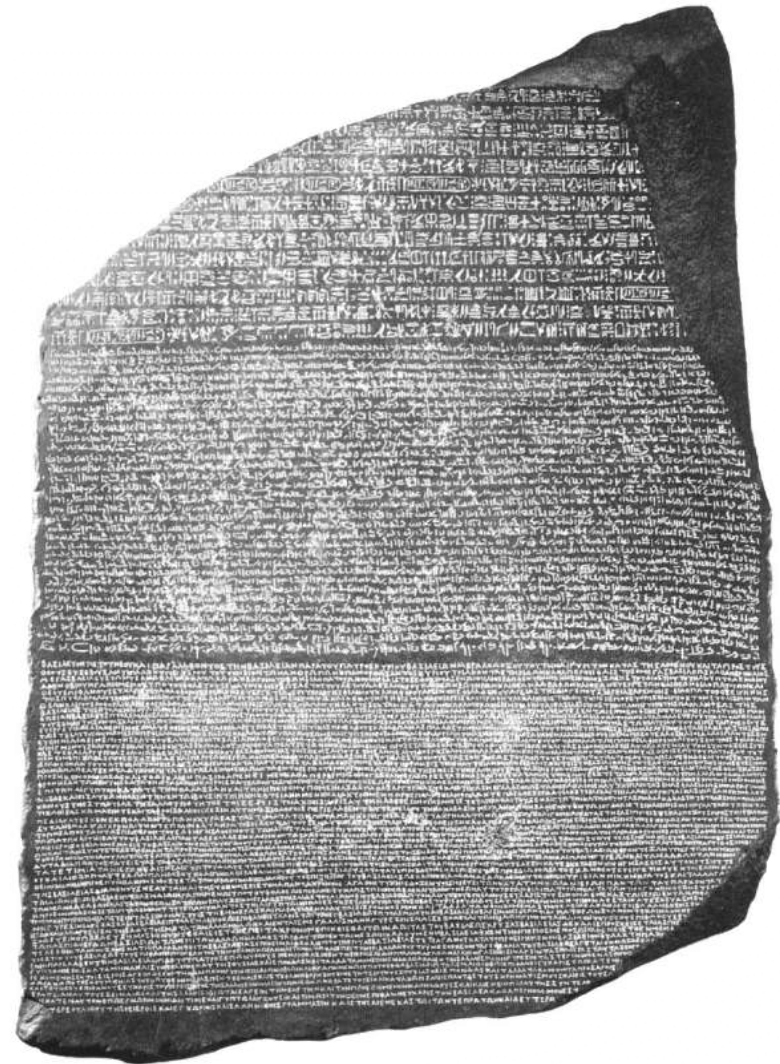
# Statistical Solution

- Parallel Texts
  - Rosetta Stone

Hieroglyphs

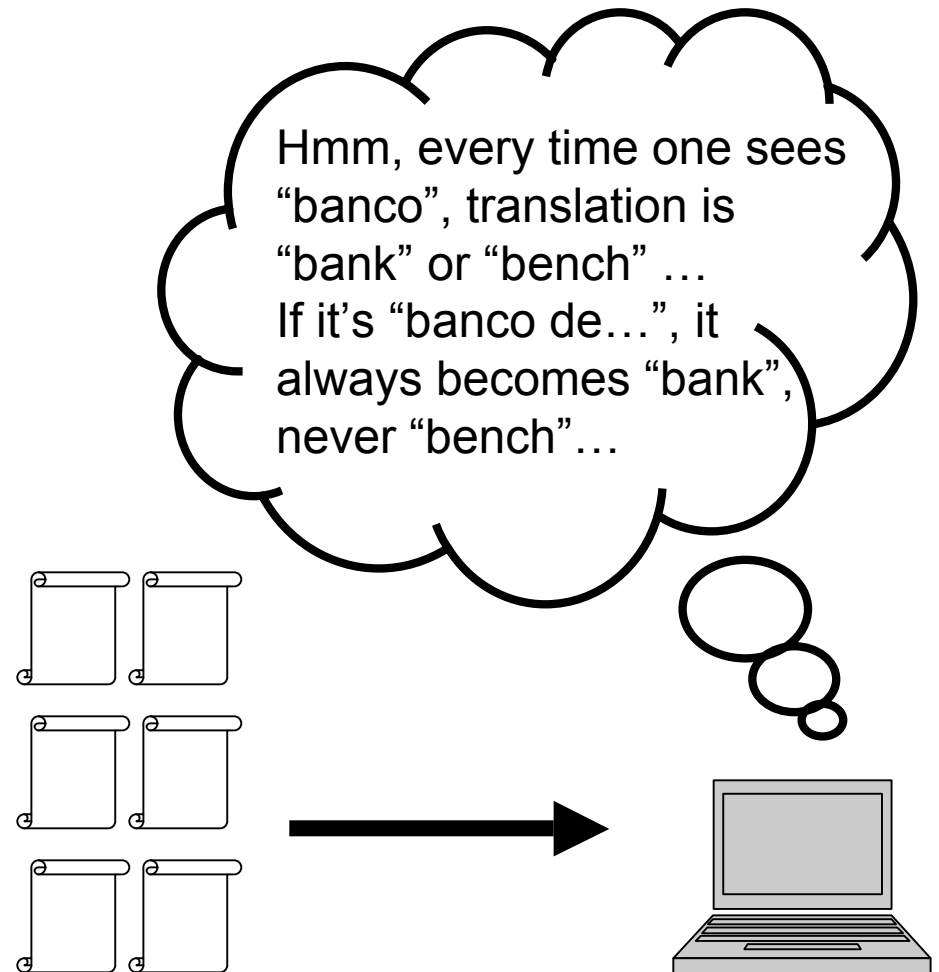
Demotic

Greek



# Statistical Solution

- Parallel Texts
  - Instruction Manuals
  - Hong Kong Legislation
  - Macao Legislation
  - Canadian Parliament Hansards
  - United Nations Reports
  - Official Journal of the European Communities



# Warren Weaver

- “Also knowing nothing official about, but having guessed and inferred considerable about, the powerful new mechanized methods in cryptography—methods which I believe succeed even when one does not know what language has been coded—one naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say: ‘This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.’ ”
  - Warren Weaver (1955:18, quoting a letter he wrote in 1947)



# Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan:    farok crrrok hihok yorok klok kantok ok-yurp

# Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: farok crrrok hihok yorok klok kantok ok-yurp

1a. ok-voon ororok sprok .	7a. lalok farok ororok lalok sprok izok enemok .
1b. at-voon bichat dat .	7b. wat jjat bichat wat dat vat eneat .
2a. ok-drubel ok-voon anak plok sprok .	8a. lalok brok anak plok nok .
2b. at-drubel at-voon pippat rrat dat .	8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghrok .	9a. wiwok nok izok kantok ok-yurp .
3b. totat dat arrat vat hilat .	9b. totat nnat quat oloat at-yurp .
4a. ok-voon anak drok brok jok .	10a. lalok mok nok yorok ghrok klok .
4b. at-voon krat pippat sat lat .	10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok .	11a. lalok nok crrrok hihok yorok zanzanok .
5b. totat jjat quat cat .	11b. wat nnat arrat mat zanzanat .
6a. lalok sprok izok jok stok .	12a. lalok rarok nok izok hihok mok .
6b. wat dat krat quat cat .	12b. wat nnat forat arrat vat gat .

# Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: **farok** crrrok hihok yorok klok kantok ok-yurp

1a. ok-voon ororok sprok .	7a. lalok <b>farok</b> ororok lalok sprok izok enemok .
1b. at-voon bichat dat .	7b. wat jjat bichat wat dat vat eneat .
2a. ok-drubel ok-voon anak plok sprok .	8a. lalok brok anak plok nok .
2b. at-drubel at-voon pippat rrat dat .	8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghrok .	9a. wiwok nok izok kantok ok-yurp .
3b. totat dat arrat vat hilat .	9b. totat nnat quat oloat at-yurp .
4a. ok-voon anak drok brok jok .	10a. lalok mok nok yorok ghrok klok .
4b. at-voon krat pippat sat lat .	10b. wat nnat gat mat bat hilat .
5a. wiwok <b>farok</b> izok stok .	11a. lalok nok crrrok hihok yorok zanzanok .
5b. totat jjat quat cat .	11b. wat nnat arrat mat zanzanat .
6a. lalok sprok izok jok stok .	12a. lalok rarok nok izok hihok mok .
6b. wat dat krat quat cat .	12b. wat nnat forat arrat vat gat .

# Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: **farok** crrrok hihok yorok klok kantok ok-yurp

1a. ok-voon ororok sprok .	7a. lalok <b>farok</b> ororok lalok sprok izok enemok .
1b. at-voon bichat dat .	7b. wat <b>jjat</b> bichat wat dat vat eneat .
2a. ok-drubel ok-voon anak plok sprok .	8a. lalok brok anak plok nok .
2b. at-drubel at-voon pippat rrat dat .	8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghrok .	9a. wiwok nok izok kantok ok-yurp .
3b. totat dat arrat vat hilat .	9b. totat nnat quat oloat at-yurp .
4a. ok-voon anak drok brok jok .	10a. lalok mok nok yorok ghrok klok .
4b. at-voon krat pippat sat lat .	10b. wat nnat gat mat bat hilat .
5a. wiwok <b>farok</b> izok stok .	11a. lalok nok crrrok hihok yorok zanzanok .
5b. totat <b>jjat</b> quat cat .	11b. wat nnat arrat mat zanzanat .
6a. lalok sprok izok jok stok .	12a. lalok rarok nok izok hihok mok .
6b. wat dat krat quat cat .	12b. wat nnat forat arrat vat gat .

# Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: **farok crrrok** hihok yorok klok kantok ok-yurp

1a. ok-voon ororok sprok .	7a. lalok <b>farok</b> ororok lalok sprok izok enemok .
1b. at-voon bichat dat .	/ 7b. wat jjat bichat wat dat vat eneat .
2a. ok-drubel ok-voon anak plok sprok .	8a. lalok brok anak plok nok .
2b. at-drubel at-voon pippat rrat dat .	8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghrok .	9a. wiwok nok izok kantok ok-yurp .
3b. totat dat arrat vat hilat .	9b. totat nnat quat oloat at-yurp .
4a. ok-voon anak drok brok jok .	10a. lalok mok nok yorok ghrok klok .
4b. at-voon krat pippat sat lat .	10b. wat nnat gat mat bat hilat .
5a. wiwok <b>farok</b> izok stok .	11a. lalok nok <b>crrrok</b> hihok yorok zanzanok .
/	???
5b. totat jjat quat cat .	11b. wat nnat arrat mat zanzanat .
6a. lalok sprok izok jok stok .	12a. lalok rarok nok izok hihok mok .
6b. wat dat krat quat cat .	12b. wat nnat forat arrat vat gat .

# Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: **farok** crrrok **hihok** yorok klok kantok ok-yurp

1a. ok-voon ororok sprok .	7a. lalok farok ororok lalok sprok izok enemok .
1b. at-voon bichat dat .	/ 7b. wat jjat bichat wat dat vat eneak .
2a. ok-drubel ok-voon anak plok sprok .	8a. lalok brok anak plok nok .
2b. at-drubel at-voon pippat rrat dat .	8b. iat lat pippat rrat nnat .
3a. erok sprok izok <b>hihok</b> ghrok .	9a. wiwok nok izok kantok ok-yurp .
3b. totat dat arrat vat hilat .	9b. totat nnat quat oloat at-yurp .
4a. ok-voon anak drok brok jok .	10a. lalok mok nok yorok ghrok klok .
4b. at-voon krat pippat sat lat .	10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok .	11a. lalok nok crrrok <b>hihok</b> yorok zanzanok .
/ 5b. totat jjat quat cat .	11b. wat nnat arrat mat zanzanat .
6a. lalok sprok izok jok stok .	12a. lalok rarok nok izok <b>hihok</b> mok .
6b. wat dat krat quat cat .	12b. wat nnat forat arrat vat gat .

# Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: **farok** crrrok **hihok** **yorok** klok kantok ok-yurp

1a. ok-voon ororok sprok .	7a. lalok farok ororok lalok sprok izok enemok .
1b. at-voon bichat dat .	7b. wat jjat bichat wat dat vat eneat .
2a. ok-drubel ok-voon anak plok sprok .	8a. lalok brok anak plok nok .
2b. at-drubel at-voon pippat rrat dat .	8b. iat lat pippat rrat nnat .
3a. erok sprok izok <b>hihok</b> ghirok .	9a. wiwok nok izok kantok ok-yurp .
3b. totat dat arrat vat hilat .	9b. totat nnat quat oloat at-yurp .
4a. ok-voon anak drok brok jok .	10a. lalok mok nok <b>yorok</b> ghirok klok .
4b. at-voon krat pippat sat lat .	10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok .	11a. lalok nok crrrok <b>hihok</b> <b>yorok</b> zanzanok .
5b. totat jjat quat cat .	11b. wat nnat arrat mat zanzanat .
6a. lalok sprok izok jok stok .	12a. lalok rarok nok izok <b>hihok</b> mok .
6b. wat dat krat quat cat .	12b. wat nnat forat arrat vat gat .

# Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: **farok** crrrok **hihok yorok** **clock** kantok ok-yurp

1a. ok-voon ororok sprok .	7a. lalok farok ororok lalok sprok izok enemok .
1b. at-voon bichat dat .	7b. wat jjat bichat wat dat vat eneat .
2a. ok-drubel ok-voon anak plok sprok .	8a. lalok brok anak plok nok .
2b. at-drubel at-voon pippat rrat dat .	8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghrok .	9a. wiwok nok izok kantok ok-yurp .
3b. totat dat arrat vat hilat .	9b. totat nnat quat oloat at-yurp .
4a. ok-voon anak drok brok jok .	10a. lalok mok nok yorok ghrok <b>clock</b> .
4b. at-voon krat pippat sat lat .	10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok .	11a. lalok nok crrrok hihok yorok zanzanok .
5b. totat jjat quat cat .	11b. wat nnat arrat mat zanzanat .
6a. lalok sprok izok jok stok .	12a. lalok rarok nok izok hihok mok .
6b. wat dat krat quat cat .	12b. wat nnat forat arrat vat gat .



# Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: **farok** crrrok **hihok yorok** **clock** kantok ok-yurp

1a. ok-voon ororok sprok .	7a. lalok farok ororok lalok sprok izok enemok .
1b. at-voon bichat dat .	7b. wat jjat bichat wat dat vat eneak .
2a. ok-drubel ok-voon anak plok sprok .	8a. lalok brok anak plok nok .
2b. at-drubel at-voon pippat rrat dat .	8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghrok .	9a. wiwok nok izok kantok ok-yurp .
3b. totat dat arrat vat hilat .	9b. totat nnat quat oloat at-yurp .
4a. ok-voon anak drok brok jok .	10a. lalok mok nok yorok ghrok <b>clock</b> .
4b. at-voon krat pippat sat lat .	10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok .	11a. lalok nok crrrok hihok yorok zanzanok .
5b. totat jjat quat cat .	11b. wat nnat arrat mat zanzanat .
6a. lalok sprok izok jok stok .	12a. lalok rarok nok izok hihok mok .
6b. wat dat krat quat cat .	12b. wat nnat forat arrat vat gat .

# Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: **farok** crrrok **hihok yorok** **clock** kantok ok-yurp

1a. ok-voon ororok sprok .	7a. lalok farok ororok lalok sprok izok enemok .   /
1b. at-voon bichat dat .	7b. wat jjat bichat wat dat vat eneak .
2a. ok-drubel ok-voon anak plok sprok .	8a. lalok brok anak plok nok . /
2b. at-drubel at-voon pippat rrat dat .	8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghrok . / /	9a. wiwok nok izok kantok ok-yurp .
3b. totat dat arrat vat hilat .	9b. totat nnat quat oloat at-yurp .
4a. ok-voon anak drok brok jok .	10a. lalok mok nok yorok ghrok clock .   / / /
4b. at-voon krat pippat sat lat .	10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok . /	11a. lalok nok crrrok hihok yorok zanzanok .   / / /
5b. totat jjat quat cat .	11b. wat nnat arrat mat zanzanat .
6a. lalok sprok izok jok stok . 	12a. lalok rarok nok izok hihok mok .   / / /
6b. wat dat krat quat cat .	12b. wat nnat forat arrat vat gat .

# Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: **farok** crrrok **hihok** **yorok** **clock** kantok ok-yurp

1a. ok-voon ororok sprok .  1b. at-voon bichat dat .	7a. lalok farok ororok lalok sprok izok enemok .   / 7b. wat jjat bichat wat dat vat eneak .
2a. ok-drubel ok-voon anak plok sprok .  2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anak plok nok . / 8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghrok . / 3b. totat dat arrat vat hilat . /	9a. wiwok nok izok kantok ok-yurp .  9b. totat nnat quat oloat at-yurp .
4a. ok-voon anak drok brok jok .  4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghrok <b>clock</b> .   X / 10b. wat nnat gat mat bat hilat . process of elimination
5a. wiwok farok izok stok . / 5b. totat jjat quat cat .	11a. lalok nok crrok hihok yorok zanzanok .   / 11b. wat nnat arrat mat zanzanat .
6a. lalok sprok izok jok stok .   6b. wat dat krat quat cat .	12a. lalok rarok nok izok hihok mok .   / 12b. wat nnat forat arrat vat gat .

# Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: **farok crrrok hihok yorok klok** kantok ok-yurp

1a. ok-voon ororok sprok .	7a. lalok farok ororok lalok sprok izok enemok .   /
1b. at-voon bichat dat .	7b. wat jjat bichat wat dat vat eneak .
2a. ok-drubel ok-voon anak plok sprok .	8a. lalok brok anak plok nok . /
2b. at-drubel at-voon pippat rrat dat .	8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghrok . / /	9a. wiwok nok izok kantok ok-yurp .
3b. totat dat arrat vat hilat .	9b. totat nnat quat oloat at-yurp .
4a. ok-voon anak drok brok jok .	10a. lalok mok nok yorok ghrok klok .   / /
4b. at-voon krat pippat sat lat .	10b. wat nnat gat mat bat hilat .   / /
5a. wiwok farok izok stok . /	11a. lalok nok crrrok hihok yorok zanzanok .   / /
5b. totat jjat quat cat .	11b. wat nnat arrat mat zanzanat .   / /
6a. lalok sprok izok jok stok . 	12a. lalok rarok nok izok hihok mok .   / /
6b. wat dat krat quat cat .	12b. wat nnat forat arrat vat gat .

cognate?

# Centauri/Arcturan [Knight, 1997]

Your assignment, put these words in order: { jjat, arrat, mat, bat, oloat, at-yurp }

1a. ok-voon ororok sprok .	7a. lalok farok ororok lalok sprok izok enemok .   /
1b. at-voon bichat dat .	7b. wat jjat bichat wat dat vat eneak .
2a. ok-drubel ok-voon anak plok sprok .	8a. lalok brok anak plok nok . /
2b. at-drubel at-voon pippat rrat dat .	8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghrok . / /	9a. wiwok nok izok kantok ok-yurp .
3b. totat dat arrat vat hilat .	9b. totat nnat quat oloat at-yurp .
4a. ok-voon anak drok brok jok .	10a. lalok mok nok yorok ghrok klok .   / / /
4b. at-voon krat pippat sat lat .	10b. wat nnat gat mat bat hilat .   / / /
5a. wiwok farok izok stok . /	11a. lalok nok <b>crrrok</b> hihok yorok zanzanak .   / <b>zero</b>
5b. totat jjat quat cat .	11b. wat nnat arrat mat zanzanat . <b>fertility</b>
6a. lalok sprok izok jok stok . 	12a. lalok rarok nok izok hihok mok .   / / /
6b. wat dat krat quat cat .	12b. wat nnat forat arrat vat gat .

# It's Really Spanish/English

**Clients do not sell pharmaceuticals in Europe => Clientes no venden medicinas en Europa**

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