Statistical Machine Translation

January 24 and 26, 2005

Dr. Andreas Eisele
Computerlinguistik & DFKI
eisele@dfki.de

Language Technology I
WS 2004/2005

Statistical Machine Translation (SMT)

UNIVERSITÄT DES SAARLANDES

- Motivation
- History and Important Players
- Mathematical Basics
- IBM Models
- Transduction Grammars
- Template-Based Approaches
- Useful Resources
- Recent Developments

Exercise:

- Learn to Translate Between Unknown Languages

Motivation for SMT

UNIVERSITÄT DES SAARLANDES

- Good translation requires knowledge and decisions on many levels
  - syntactic disambiguation (POS, attachments)
  - semantic disambiguation (collocations, scope, word sense)
  - reference resolution
  - lexical choice in target language
  - application-specific terminology, register, connotations, good style …
- Rule-based models of all these levels are very expensive to build, maintain, and adapt to new domains
- Statistical approaches have been quite successful in many areas of NLP, once data has been annotated
- Learning from existing translation will focus on distinctions that matter (not on the linguist's favorite subject)
- Translation corpora are available in rapidly growing amounts
- SMT *can* integrate rule-based modules (morphologies, lexicons)
- SMT *can* use feed-back for on-line adaptation to domain and user preferences

History and Important Players

UNIVERSITÄT DES SAARLANDES

- 1949: Warren Weaver: *the translation problem can be largely solved by "statistical semantic studies"*
- 1950s..1970s: Predominance of rule-based approaches
- 1966: ALPAC report: general discouragement for MT (in the US)
- 1980s: example-based MT proposed in Japan (Nagao), statistical approaches to speech recognition (Jelinek e.a. at IBM)
- Late 80s: Statistical POS taggers, SMT models at IBM, work on translation alignment at Xerox (M. Kay)
- 90s: many statistical approaches to NLP in general, IBM's Candide claimed to be as good as Systran
- Late 90s: Statistical MT successful as a fallback approach within Verbmobil System (Ney, Och). Wide distribution of translation memory technology (Trados) indicates big commercial potential of SMT
- 1999 Johns Hopkins workshop: OS re-implementation of IBM methods
- since 2001: DARPA evaluation campaign (XYZ -> English)
- Various companies marketing/exploring SMT technologies:
  language weaver, aixplain GmbH, Linear B Ltd., esteam, Google Labs

Related Fields

UNIVERSITÄT DES SAARLANDES

- Initially, statistical and knowledge-based approaches were seen in strict contrast.
- However, PROs and CONs are somewhat complementary:

	Syntax	Structural Semantics	Lexical Semantics	Adaptivity
Rule-based MT	++	+	-	--
Statistical MT	--	--	+	+
Example-based MT	-	--	-	++

➢ It is now more or less consensus to target integrated approaches

Hybrid Approaches

UNIVERSITÄT DES SAARLANDES

In the beginning...

	SMT	EBMT	Transfer
Hypothesis probabilities	1		
Collocation bilexicons		1	
Transduction rules			1

HKUST Human Language Technology Center   Dekai Wu, MT-Summit 2003, 2003.09.26

Hybrid Approaches

But then…

- SMT plants trees
  - got collocation bilexicons?
    - learning: Wu & Xia 1995, Smadja 1996, Och et al. 1999, Koehn et al. 2003
    - decoding: Wu 1996, Och et al. 1999, Koehn et al. 2003
- EBMT gets serious about template abstraction
  - got transduction rules?
- Transfer models string out
  - got collocation bilexicons?

HKUST Human Language Technology Center　　　Dekai Wu, MT-Summit 2003, 2003.09.26

So then…

	SMT	EBMT	Transfer
Hypothesis probabilities	0.6		
Collocation bilexicons	0.4	0.6	0.4
Transduction rules		0.4	0.6

HKUST Human Language Technology Center Dekai Wu, MT-Summit 2003, 2003.09.26

And then…

- SMT plants trees
  - got collocation bilexicons?
    - learning: Wu & Xia 1995, Smadja 1996, Och et al. 1999, Koehn et al. 2003
    - decoding: Wu 1996, Och et al. 1999, Koehn et al. 2003
  - got 'real linguistic' transduction grammar rules?
    - eg: Wu & Wong 1998, Alshawi et al. 1998, Yamada & Knight 2001, Melamed 2003, Schafer & Yarowsky 2003
- EBMT gets real about scoring
  - got probabilities?
    - eg: Brown et al. 2003
- Transfer models soften up
  - got scores, backoff, stronger decoders?
    - eg: Lavie et al. 2003

Hybrid Approaches

So then...

	SMT	EBMT	Transfer
Hypothesis probabilities	0.5	0.2	0.2
Collocation bilexicons	0.3	0.5	0.3
Transduction rules	0.2	0.3	0.5

HKUST Human Language Technology Center — Dekai Wu, MT-Summit 2003, 2003.09.26

Hybrid Approaches

UNIVERSITÄT DES SAARLANDES

Convergence

	SMT	EBMT	Transfer
Hypothesis probabilities	0.33	0.33	0.33
Collocation bilexicons	0.33	0.33	0.33
Transduction rules	0.33	0.33	0.33

HKUST Human Language Technology Center Dekai Wu, MT-Summit 2003, 2003.09.26

But:

before integrating things, we need to know them in detail!

Mathematical Basics

UNIVERSITÄT DES SAARLANDES

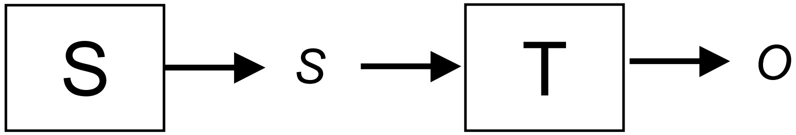
Assume the probabilities of English and French sentences e and f appearing as translations.

P(e): prior probability of e, w/o additional knowledge
P(f|e): conditional probability of f, given e
P(f,e) = P(e) * P(f|e): joint probability
P(f|e) = P(f) * P(e|f) / P(e): Bayes' Rule

Question: P(f,e) = P(f) * ??

"Distorted Channel" Paradigm

Assume a signal that has to be transmitted through a channel that may add distortion, noise, or other modifications:



Both the source of the signal and the transmission channel can be characterized as probability distributions:

P(s): propability that signal s is generated

P(o|s): probability that observation o is made, *given* s

P(o,s) = P(s) * P(o|s): probability that s is sent *and* o is observed

In typical applications, the most likely cause s' for a given observation o is sought, i.e.

$$s' = \text{argmax}_s\ P(s|o) = \text{argmax}_s\ P(s,o) = \text{argmax}_s\ P(s) * P(o|s)$$

Applications of Distorted Channel Paradigm

UNIVERSITÄT DES SAARLANDES

Communications Engineering:

- S may be an input device, T a transmission line (modem line, audio/video transmission)

Speech recognition:

- S is the speaker's brain, generating a string of words
- T is the chain consisting of speakers articulatory device, sound transmission, microphone, signal processing up to morpheme hypotheses. The task is to reconstruct from a string of decoded sound events the intended chain of words.

Machine translation:

- S is text in one language, T is translation to another
- applying this model means to translate from the target language of the assumed "distortion" to the source, which can be confusing

Error correction

- S is the intended (correct) text, T is the modification by introducing typing, spelling and other errors

OCR, …

Important Properties of this Model

UNIVERSITÄT DES SAARLANDES

- P(S) and P(O|S) can be modeled independently

- P(S) can be approximated using large amounts of monolingual text (e.g. using n-gram models)
- The same monolingual model P(S) can be useful for ASR, OCR, and SMT
- Models based on grammars and other knowledge sources are harder to build, but will be superior in the long run

- P(O|S) can be approximated from translated corpora, when correspondences between source and target language are known or can be estimated
- Again, models that take linguistic structure into account will be better in the long run, but require more effort to be built

IBM Translation Models

UNIVERSITÄT DES SAARLANDES

- Brown e.a. 1993 propose 5 different ways to define P(f|e) and to train the parameters from a bilingual corpus
- There is a chicken-and-egg situation between translation models and alignments: given one, we can estimate the other. The standard approach to bootstrap reasonable models from partially hidden data is the Expectation-Maximization (EM-) Algorithm (as also used e.g. for HMMs)
- Model 1 assumes a one-to-one relation between individual words and a uniform distribution over all possible permutations
- Model 2 is similar, but prefers alignments that roughly preserve the original order

Word Alignment Example from Europarl

UNIVERSITÄT DES SAARLANDES

	Frau	Ludford	,	möchten	Sie	auch	wirklich	eine	Anmerkung	zum	Protokoll	machen	?
NULL	.	.	.	.	.	*	.	*	.	.	.	*	.
Mrs	*	.	.	.	.	.	.	.	.	.	.	.	.
Ludford	.	*	.	.	.	.	.	.	.	.	.	.	.
,	.	.	*	.	.	.	.	.	.	.	.	.	.
are	.	.	.	*	.	.	.	.	.	.	.	.	.
you	.	.	.	.	*	.	.	.	.	.	.	.	.
sure	.	.	.	.	.	.	*	.	.	.	.	.	.
your	.	.	.	.	.	.	.	.	.	.	.	.	.
point	.	.	.	.	.	.	.	.	*	.	.	.	.
of	.	.	.	.	.	.	.	.	.	.	.	.	.
order	.	.	.	.	.	.	.	.	.	.	.	.	.
is	.	.	.	.	.	.	.	.	.	.	.	.	.
related	.	.	.	.	.	.	.	.	.	.	.	.	.
to	.	.	.	.	.	.	.	.	.	.	.	.	.
the	.	.	.	.	.	.	.	.	.	*	.	.	.
Minutes	.	.	.	.	.	.	.	.	.	.	*	.	.
?	.	.	.	.	.	.	.	.	.	.	.	.	*

IBM Translation Models

UNIVERSITÄT DES SAARLANDES

- Model 3 assumes that one English word can give rise to multiple French words by introducing "fertilities", i.e. distributions over the number of words in the translation of a given word. Exact calculation of EM-estimates becomes infeasible and is replaced with approximations restricted to plausible subsets of all possible alignments.
- Model 4 introduces a distinction between groups of words (derived from one source word) that tend to stay together (like: *implemented* → *mis en application*) and groups that tend to get separated (like: not → ne … pas).
- Model 5 is similar to Model 4, but avoids to distribute probability mass over impossible word sequences, e.g. sequences where words are missing or positions are simultaneously occupied with more than one word.
- Formulas in the CL'93 paper look heavy, but there are many tutorials and even an open-source implementation available.

IBM Translation Models

- Bootstrapping also works across models of increasing complexity (i.e. alignment from Model *i* is used to estimate parameters for Model *i+1*)
- Development of the IBM models was based on about 1.8 million sentence pairs from the Canadian parliament debates (Hansards)
- Decoding (i.e. search for argmax_s P(s) * P(o|s)) was computationally challenging for long sentences, hence various heuristics for sentence splitting were used
- The performance was evaluated in a ’94 ARPA test; Candide translations were judged as more fluent, but less adequate than those of Systran
- All models assume that correspondences are triggered by single words on the source level side, i.e. there is no support for phrase-to-phrase alignments

Inversion Transduction Grammars

UNIVERSITÄT DES SAARLANDES

- Instead of a directed translation relation between input and output strings, transduction grammars [Wu 199x…] model the simultaneous stochastic generation of strings in two languages.
- Differences in word order can be explained by rules like

  NP → [Adj N]

  generating *Adj N* on one side and *N Adj* on the other
- Parsing sentence pairs using this type of grammars can be done by generalizing context-free parsing to two dimensions (~ squaring the computational complexity)
- Optimal decoding is possible in polynomial time, e.g. $O(n^7)$

Template-Based Approaches

- Models used by Ney/Och in Verbmobil and later at ISI are inspired by, but different from the IBM models
- They are based on the alignment of “phrases” in both languages (phrase = any sequence of words)
- Translation segments the input to phrases, translates each of the phrases, and re-orders the outcomes in the target language
- Search for best translation searches for optimal combined solution for all of these steps, including source language segmentation (= some very limited form of parsing)

Recent Developments

UNIVERSITÄT DES SAARLANDES

- Since 2001: DARPA organizes a evaluation campaign for translation from Chinese, Arabic, and more languages to English. For some tests, the source language was known only a couple of weeks before the submission deadline. Practically all systems use SMT
- Philip Koehn (ISI, now MIT) collected and published a 11-language Corpus of European Parliament debates (20..30 MW/lang)
- OPUS project collected and aligned large number of texts (e.g. documentation of open-source software)
- Johns Hopkins Workshop 2003 explored “Syntax for Statistical Machine Translation” and reported some improvements via the inclusion of syntactic features
- Google Labs hired a group of world-class researchers (including Torsten Brants, Franz Och) to work on SMT

Some entry points…

UNIVERSITÄT DES SAARLANDES

- Valuable Material (SW, corpora, papers) can be found at homepages of the following persons:
  - Franz Och (former homepage at Aachen)
  - Kevin Knight (ISI)
  - Philipp Koehn (MIT)
  - Andy Way (Dublin)

Centauri/Arcturan [Knight, 1997]

Your assignment, translate this to Arcturan: farok crrrok hihok yorok clok 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 eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghirok clok . 10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok . 5b. totat jjat quat cat .	11a. lalok nok crrrok 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 clok 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 eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghirok clok . 10b. wat nnat gat mat bat hilat .
5a. wiwok **farok** izok stok . 5b. totat jjat quat cat .	11a. lalok nok crrrok 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 clok 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 eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghirok clok . 10b. wat nnat gat mat bat hilat .
5a. wiwok **farok** izok stok . 5b. totat **jjat** quat cat .	11a. lalok nok crrrok 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 clok 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 eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghirok clok . 10b. wat nnat gat mat bat hilat .
5a. wiwok **farok** izok stok . 5b. totat jjat quat cat .	11a. lalok nok **crrrok** 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 clok 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 eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok **hihok** ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghirok clok . 10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok . 5b. totat jjat quat cat .	11a. lalok nok crrrok **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 clok 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 eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok **hihok** ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok **yorok** ghirok clok . 10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok . 5b. totat jjat quat cat .	11a. lalok nok crrrok **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** clok 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 eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghirok **clok** . 10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok . 5b. totat jjat quat cat .	11a. lalok nok crrrok 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]

UNIVERSITÄT DES SAARLANDES

Your assignment, translate this to Arcturan: **farok** crrrok **hihok yorok** clok 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 eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghirok **clok** . ??? 10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok . 5b. totat jjat quat cat .	11a. lalok nok crrrok 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** clok 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 eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghirok clok . 10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok . 5b. totat jjat quat cat .	11a. lalok nok crrrok 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]

UNIVERSITÄT DES SAARLANDES

Your assignment, translate this to Arcturan: **farok** crrrok **hihok yorok clok** 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 eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghirok clok . 10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok . 5b. totat jjat quat cat .	11a. lalok nok crrrok 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 .

process of elimination

Centauri/Arcturan [Knight, 1997]

UNIVERSITÄT DES SAARLANDES

Your assignment, translate this to Arcturan: **farok** crrrok **hihok yorok clok** 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 eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghirok clok . 10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok . 5b. totat jjat quat cat .	11a. lalok nok crrrok hihok yorok zanzanok . 11b. wat nnat arrat mat zanzanat . cognate?
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]

UNIVERSITÄT DES SAARLANDES

Your assignment, put these words in order: **{ jjat, arrat, mat, bat, oloat, at-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 eneat .
2a. ok-drubel ok-voon anok plok sprok . 2b. at-drubel at-voon pippat rrat dat .	8a. lalok brok anok plok nok . 8b. iat lat pippat rrat nnat .
3a. erok sprok izok hihok ghirok . 3b. totat dat arrat vat hilat .	9a. wiwok nok izok kantok ok-yurp . 9b. totat nnat quat oloat at-yurp .
4a. ok-voon anok drok brok jok . 4b. at-voon krat pippat sat lat .	10a. lalok mok nok yorok ghirok clok . 10b. wat nnat gat mat bat hilat .
5a. wiwok farok izok stok . 5b. totat jjat quat cat .	11a. lalok nok crrrok hihok yorok zanzanok . 11b. wat nnat arrat mat zanzanat . zero fertility
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 .

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 .