Entropy

Leonoor van der Beek, vdbeek@rug.nl
Department of Alfa-informatica
Rijksuniversiteit Groningen

May 2005

RuG

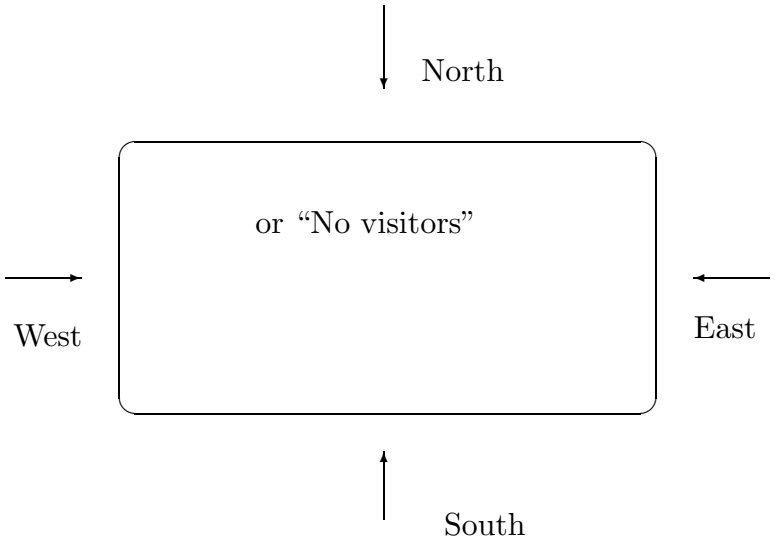
What is entropy?

Entropy is a measure of uncertainty or surprise or disorder.

Entropy was first applied in physics (disorder of gas) and then in telecommunication (optimal, most compact coding)

Entropy, an example

Consider the situation where a lookout must report either no visitor or the direction from which a visitor is approaching, i.e. one of five messages:



Should we code 000, 001, 010, 011, 100? Then all codes have three bits.

Example: code length

With no further information, we seem to need a code length

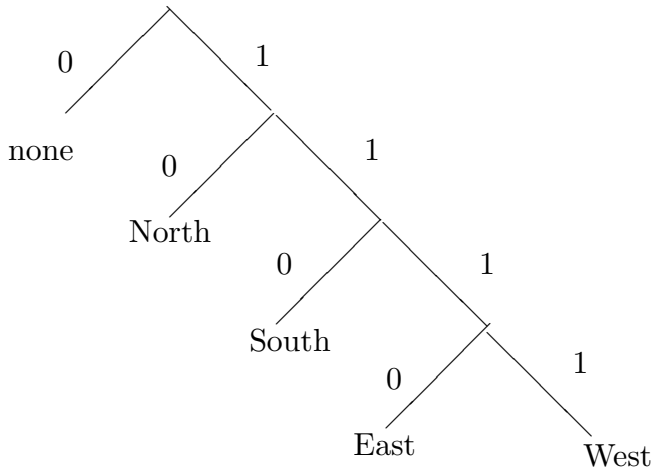
$$\text{code length} = \lceil \log_2 |M| \rceil, \text{ where } M \text{ are the messages}$$

But suppose we know that some messages are more frequent than others. . .

message	rel. freq.
no visitor	99%
North	0.5%
South	0.25%
East, West	0.125%

RuG

Example: a code tree



message	code
no visitor	0
North	10
South	110
East	1110
West	1111

RuG

Example: expected Code Length

We now calculate the expected code length:

message	code length	rel. freq.	expected bit length
no visitor	1	0.99	0.99
North	2	0.005	0.01
South	3	0.0025	0.0075
East	4	0.00125	0.005
West	4	0.00125	0.005
Total			1.0175

Compare to 3 bits,

$$\text{code length} = \lceil \log_2 |M| \rceil, \text{ where } M \text{ are the messages}$$

RuG

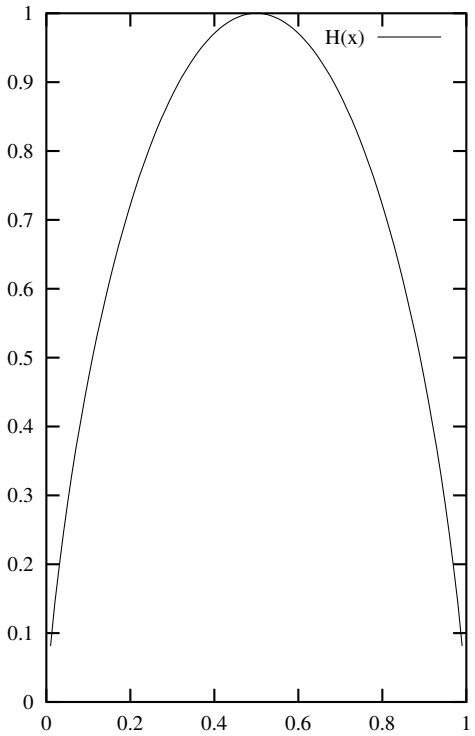
Entropy

The optimal code cannot be compressed further than the **entropy** (informational uncertainty) of the dataset:

$$H(S) = -\sum_{i \in S} p_i \log_2 p_i$$

message	p_i	$-\log p_i$	$p_i \log p_i$
no visitor	0.99	0.004	0.0044
North	0.005	2.3	0.0115
South	0.0025	2.6	0.0065
East	0.00125	2.9	0.0036
West	0.00125	2.9	0.0036
Total			0.021

Entropy of Two-Way Choice



RuG

Entropy reduction

By adding knowledge to the system, one reduces the uncertainty. The information gain can be quantified by comparing the total entropies of the original system and the final system.

Suppose visitors never come on Mondays. Then adding information about the day of the week will reduce the entropy:

Day	P	Entropy
Mondays	0.143	0
Other	0.857	0.021
Total		0.018

Entropy in linguistics?

- In language, variation is often possible: active vs. passive, topicalization vs. in situ, synonym pairs, NP vs. PP recipient.
- Which variant is realized is often influenced by various factors.
- The influence of a factor can be quantified by means of entropy reduction.

The dative alternation in English

(1) a. Kim gave Bo the book.
b. Kim gave the book to Bo.

Both the syntactic category of the recipient and the order of the two complements change.

Analyses of this alternation have been formulated in terms of general alignment constraints, verbal preferences and feature/category alignment.

The dative alternations in Dutch

(2) Kim gaf Bo het boek.
Kim gave Bo the book
Kim gave Bo the book

(3) Kim gaf het Bo.
Kim gave it Bo
Kim gave it to Bo.

(4) Kim gaf het book aan Bo.
Kim gave the book to Bo
Kim gave the book to Bo.

(5) Kim gaf aan Bo het boek.
Kim gave to Bo the book
Kim gave the book to Bo.

The order of the complements and the syntactic category of the recipient vary independently.

Hypothesis (I)

Given that argument order and recipient category alternate independently, we expect general alignment constraints to influence the ordering, and lexical preferences/harmonic alignment constraints to influence the NP/PP alternation.

Hypothesis (II)

- The category of the direct object influences argument order
- The verb lexeme influences the NP/PP alternation

Illustration

(6) a. Kim geeft jou het boek/?dat/?het
Kim gives you the book/that/it

b. Kim geeft *het boek/dat/het jou
Kim gives the book/that/it you

(7) a. Ik vraag/?rapporteer/verwijt Kim dat
I ask/report/reproach Kim that

b. Ik vraag/rapporteer/??verwijt dat aan Kim
I ask/report/reproach that to Kim

Log-Likelyhood

As a first indication, we test whether the association between verb lexeme and order/recipient category is significant.

Since we already know that the category of the direct object influences argument order, we test this for each category separately.

Log-Likelyhood

Alternation	Obj1	Degrees of Freedom	LL	Significant
Arg Order (NP NP)	NP	35	6.2	no
	pron	20	22.9	no
	het	7	4.4	no
NP/PP Alternation	NP	40	79.8	p=0.001
	pron	24	36.5	p=0.050
	het	7	8.3	no

Can we quantify this influence and compare it to the influence of obj1 category?

Entropy

Alternation	Ent before	Ent Cat		Ent Verb		Ent Cat+Verb	
NPNP order	0.172	0.110	-36%	0.152	-12%	0.094	-45%
NP/PP alternation	0.578	0.578	-0%	0.426	-26%	0.422	-27%

Entropy

Alternation	Ent before	Ent Cat		Ent Verb		Ent Cat+Verb	
NPNP order	0.172	0.110	-36%	0.152	-12%	0.094	-45%
NP/PP alternation	0.578	0.578	-0%	0.426	-26%	0.422	-27%

- With less categories, category information gives a much greater entropy reduction for arg order than verb lexeme.
- Obj1 category information does not reduce the entropy for NP/PP alternation; verb lexeme information does.

Conclusion

Entropy reduction provides a means of quantifying the influence of a particular factor on a particular linguistic alternation.