NORTH ATLANTIC TREATY
ORGANIZATION

SCIENCE AND TECHNOLOGY
ORGANIZATION

AC/323()

www.sto.nato.int

ANNEX A

I think it is possible it might be so...Using Lexical Clues to Generate Evidence Weights

Kellyn Rein

I think it is possible it might be so...
Using Lexical Clues to Generate Evidence Weights

Kellyn Rein

Fraunhofer FKIE

Command and Control Systems

Information Analysis

Fraunhoferstr. 20, 53343 Wachtberg

GERMANY

kellyn.rein@fkie.fraunhofer.de

We are drowning in information but starved for knowledge.

John Naisbitt

"According to computer giant IBM, 2.5 exabytes - that's 2.5 billion gigabytes (GB) - of data was generated every day in 2012."

"Think of it this way—five exabytes of content were created between the birth of the world and 2003. *In 2013, 5 exabytes of content were created each day*." (italics added)

Volume of data produced every *minute* of every day (by early 2014!)
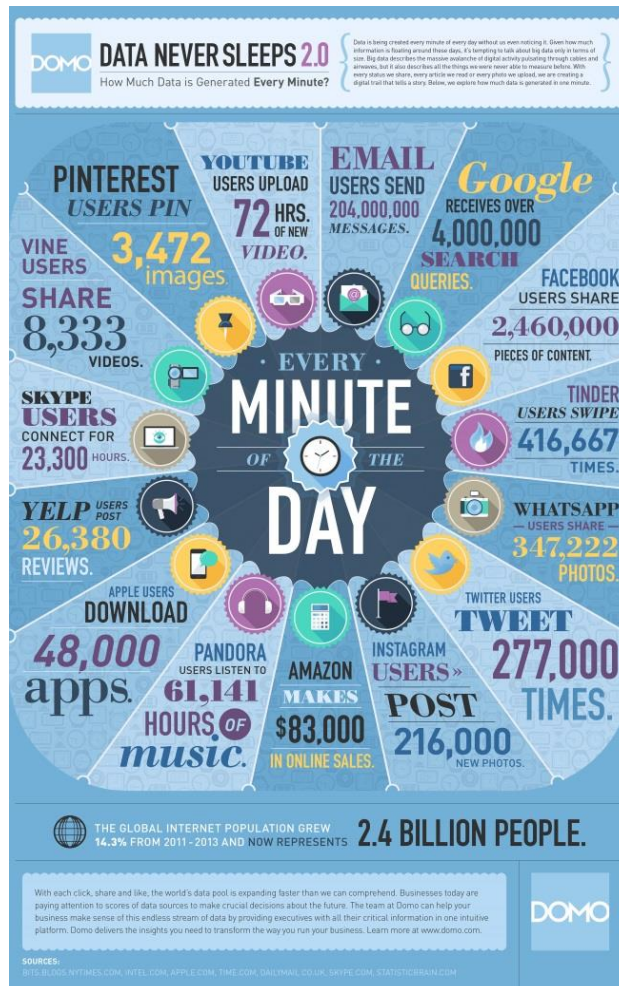
SAS-114, Copenhagen, Dec 2016

There are two basic categories of detectable uncertainty which appear at the sentence level within written text or in speech:

Uncertainty **within** the content, including
- Imprecision
- Vagueness
- Ambiguity and polysemy (multiple meanings)

Uncertainty **about** the content, including:
- Modal verbs
- Modal adverbs (including "words of estimative probability")
- Hearsay markers
- "Mindsay" markers → belief, inference, assumption, etc.
- Passive voice

The uncertainty **about** the data is content uncertainty (data level),while uncertainty **within** the content comes into play at the correlation of various discrete elements of data (fusion level).

There are two basic categories of detectable uncertainty which appear at the sentence level within written text or in speech:

Uncertainty **within** the content, including
>> Imprecision
>> Vagueness
>> Ambiguity and polysemy (multiple meanings)

Uncertainty **about** the content, including:
>> Modal verbs
>> Modal adverbs (including "words of estimative probability")
>> Hearsay markers
>> "Mindsay" markers → belief, inference, assumption, etc.
>> Passive voice

The uncertainty **about** the data is content uncertainty (data level),while uncertainty **within** the content comes into play at the correlation of various discrete elements of data (fusion level).

Raison d'etre

- In numerous application areas information extracted from natural language text is being used for decision-making

- Up to now, assessment of the quality of such extracted information has been lacking
  - when (if) extracted information is evaluated it is generally done by a human and done out of its original context

- However, there is lexical information embedded in the text which give clues as to origin and credibility of the information contained in the proposition which can be exploited for automatic evaluation and assignment of weights

Uncertainty *about* the content:
Linguistic markers of uncertainty

Specific content within a given statement is often packed with lexical elements that indicate in some manner the uncertainty of the content itself or that indicate the original source of information.

- *John is a terrorist.*
- *The CIA have concluded that John is a terrorist.*
- *I believe that John is a terrorist.*
- *My neighbor thinks John is a terrorist.*
- *It has been definitely disproved that John is a terrorist.*
- *Unless things change, John will be a terrorist one day.*
- *The CIA have concluded that John is probably a terrorist.*
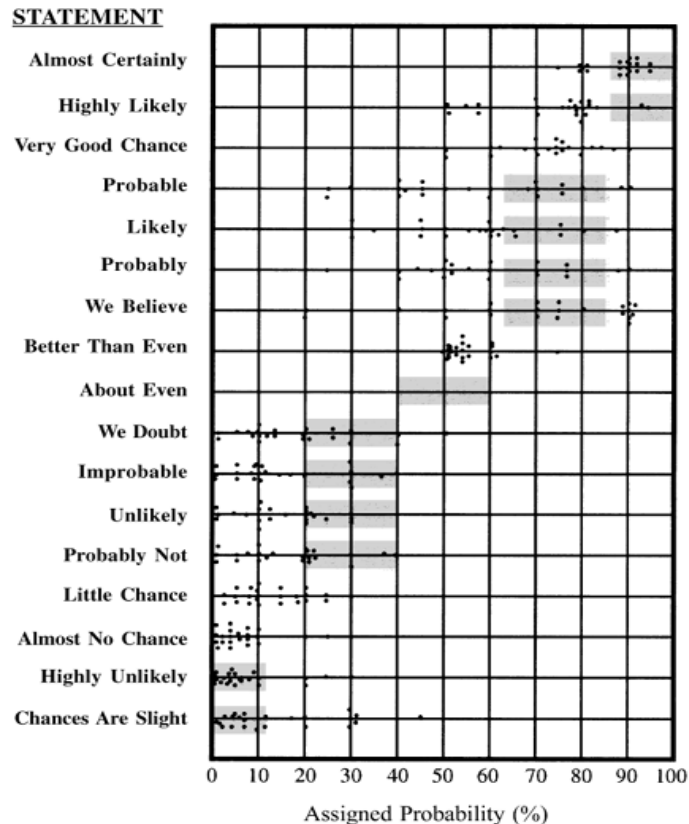
Uncertainty *about* the content:
Linguistic markers of uncertainty

"A few days after the estimate ["NIE 29-51, "Probability of an Invasion of Yugoslavia in 1951"] appeared, I was in informal conversation with the Policy Planning Staff's chairman. We spoke of Yugoslavia and the estimate. Suddenly he said, "By the way, what did you people mean by the expression `**serious possibility'**? …I told him that my personal estimate was on the dark side, namely, that the **odds were around 65 to 35 in favor of an** attack. He was somewhat jolted by this; he and his colleagues had read "**serious possibility" to mean odds very considerably lower.**

"[it turned out that] each Board member had had somewhat different odds in mind and the low man was thinking of about 20 to 80, the high of 80 to 20. The rest ranged in between.

Uncertainty *about* the content:
Linguistic markers of uncertainty

Figure 18: Measuring Perceptions of Uncertainty

Results of CIA analyst Sherman Kent's informal study of weighting by colleagues

Uncertainty *about* the content:
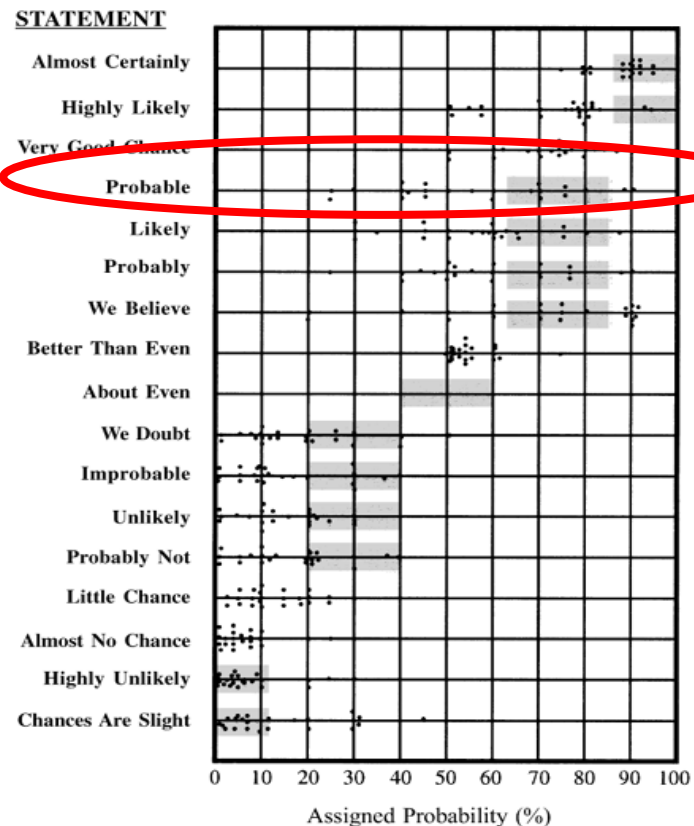# Linguistic markers of uncertainty

Figure 18: Measuring Perceptions of Uncertainty

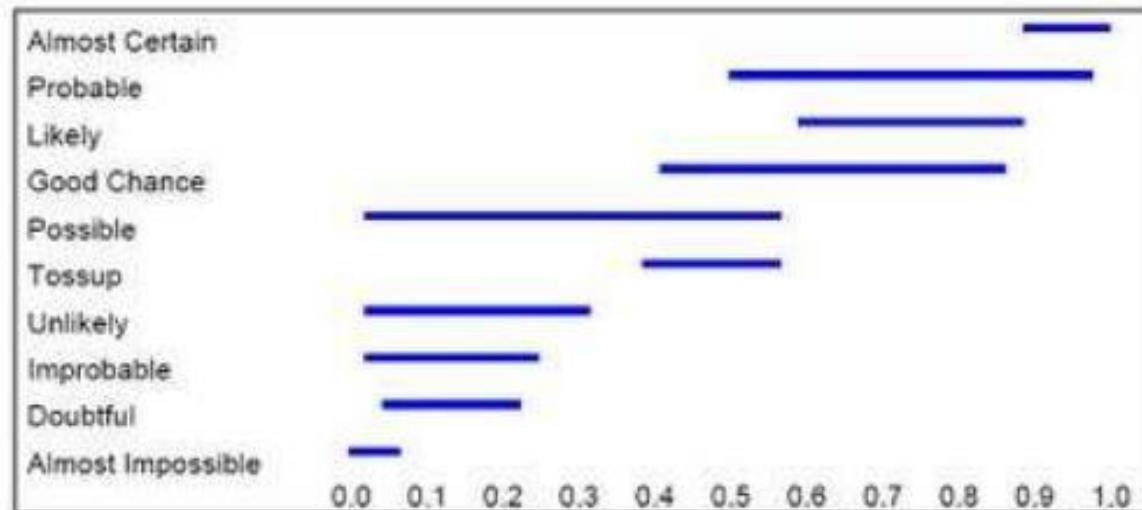Results of CIA analyst Sherman Kent's informal study of weighting by colleagues

Uncertainty *about* the content:
Linguistic markers of uncertainty

Ranges of percentages assigned to hedges by analysts in training.

Uncertainty *about* the content:
Linguistic markers of uncertainty

Words of Estimative Probability, as appeared in the 2007 National Intelligence Estimate, Iran Nuclear Intentions and Capabilities as well as in the front matter of several other recent intelligence products

| Remote | Very unlikely | Unlikely | Even chance | Probably/ Likely | Very likely | Almost certainly |

Uncertainty *about* the content:
Linguistic markers of uncertainty

Specific content within a given statement is often packed with lexical elements that indicate in some manner the uncertainty of the content itself or that indicate the original source of information.

- *John is a terrorist.*
- *The CIA have concluded that John is a terrorist.*
- *I believe that John is a terrorist.*
- *My neighbor thinks John is a terrorist.*
- *It has been definitely disproved that John is a terrorist.*
- *Unless things change, John will be a terrorist one day.*
- *The CIA have concluded that John is probably a terrorist.*

Uncertainty *about* the content:
Linguistic markers of uncertainty

Hedges and evidential (how the information was acquired) markers are relatively obvious indicators of uncertainty, even to non-linguists. However, there are some more subtle ways in which uncertainty may appear.

…the writer inevitably uses a wide range of depersonalized forms which shift responsibility for the validity of what is asserted from the writer to those whose views are being reported. Verb forms such as *argue*, *claim*, *contend*, *estimate*, *maintain* and *suggest* occurring with third person subjects are typical examples of forms functioning in the way, as are adverbials like *allegedly*, *reportedly*, *supposedly* and *presumably*.

Uncertainty *about* the content:
Linguistic markers of uncertainty

Passive voice:

- Particularly in scientific writing, the use of passive voice and impersonal phrasing are widely,
  "I might be wrong or have overlooked something."

- Can also be used to express politeness, rather than uncertainty, which can only be determined by knowing some information about the context of the statement.

- Also sometimes used in the case of differences in social ranking or power, in order not to offend – again more information needed

Uncertainty *about* the content:
Linguistic markers of uncertainty

Verb tenses / moods / temporal expressions

- Future is inherently uncertain because *it may not happen*
    That being said, some future things are more certain than others:
        "The next presidential election in the US will take place in November 2016."

- Expressions of routine activity: *The group meets every Monday at 10 a.m.*
    ("Well, not next week because it's a holiday…")

For intelligence purposes, information based upon future actions often plays a very significant role, but should nearly always be considered uncertain, until the expected date of that action has passed (and it has or has not occurred).

Uncertainty *about* the content:
Linguistic markers of uncertainty

PROBABILITY RATINGS (0 TO 6 SCALE) AND PERCEIVED AMBIGUITY OF PROBABILITY WORDS IN A MEDICAL TREATMENT CONTEXT (STUDY II)

| | No context (students) | | Treatment context | | | | | |
| | | | Physicians | | | Parents | | |
Expression	Mean	SD	Mean	SD	Perceived ambiguity (%)	Mean	SD	Perceived ambiguity (%)
Impossible	0.1	0.4	0.0	0.2	96	0.1	0.2	93
Improbable	1.2	1.1	0.6	0.5	87	0.3	0.5	83
Doubtful	1.3	0.7	1.0	0.4	92	0.9	0.5	88
Perhaps	2.9	0.8	2.3	0.9	73	2.4	0.9	73
Possibly	3.1	1.0	2.2	0.9	73	2.5	1.1	70
Chance for	3.1	0.9	2.3	1.0	77	2.7	0.9	67
Danger for	3.3	1.3	1.5	0.9	76	1.1	1.0	75
Possible	3.6	0.8	2.5	0.9	79	2.9	0.9	74
Assumedly	4.1	1.0	3.4	1.1	65	3.5	1.2	69
Good hope	4.2	0.9	3.9	0.9	84	4.4	1.0	77
Likely	4.2	0.9	3.9	0.9	76	3.5	1.0	69
Good chance	4.5	0.6	4.2	0.8	86	4.5	0.9	79
Probable	4.6	0.8	4.4	0.8	73	3.8	1.1	66
Small doubt	4.6	1.2	4.8	0.9	64	4.1	1.9	54
Mean	3.2	0.9	2.6	0.8	79	2.6	0.9	74

Weights assigned to expressions of uncertainty used in the context of medical discussions between pediatricians and the parents of the sick children. Brun- and Teigen [bru:88, p.397]

Uncertainty *about* the content:
Linguistic markers of uncertainty

PROBABILITY RATINGS (0 TO 6 SCALE) OF PROBABILITY WORDS USED IN A VARIETY OF
CONTEXTS (TV NEWS REPORTS)

Expression	Within context		Without context	
	Mean	SD	Mean	SD
Impossible	0.6	0.7	0.1	0.4
	0.8	1.0		
Improbable	0.6	0.6	1.2	1.1
Small probability	1.3	0.9	—	—
	1.4	1.1		
Perhaps*	3.2	1.0	2.9	0.8
	3.4	1.4		
	3.9	1.3		
	4.6	0.8		
	5.0	0.7		
Possibly*	3.6	1.8	3.1	1.0
	3.6	1.2		
	4.5	1.1		
Possible*	4.2	1.3	3.6	0.8
	4.3	1.1		
	5.0	1.3		
Assumedly	4.3	0.8	4.1	1.0
	4.3	1.1		
	4.8	1.0		
Likely*	2.1	1.8	4.2	0.9
	4.5	1.5		
	4.7	0.9		
	5.0	0.7		
	5.0	0.8		
Good chance	3.8	1.2	4.5	0.6
Probable*	4.5	1.1	4.6	0.8
	5.1	0.7		

Note. Asterisks indicate expressions with significantly different mean scores ($p < .05$) in different contexts.

Weights assigned to probabilistic expressions used in televised news reports. Brun-and Teigen [bru:88, p. 401]

Uncertainty *about* the content:
Linguistic markers of uncertainty

PROBABILITY ESTIMATES AND PERCEIVED AMBIGUITY OF PROBABILITY PHRASES UNDER THREE DIFFERENT CONDITIONS (STUDY I)

Expression	Group I			Group II		Group III (context)		
	Mean	SD	Perceived ambiguity (%)	Mean	SD	Mean	SD	r with own opinion
Impossible	.02	.03	90	.00	.00	.07	.07	.07
Not possible	.03	.03	81					
No chance	.03	.03	87					
Improbable	.08	.06	63	.12	.06	.12	.11	−.01
Very doubtful	.09	.06	67					
Small possibility	.13	.07	72					
Not probable	.13	.09	55					
Small probability	.13	.07	68					
Not likely	.13	.10	47					
Quite doubtful	.15	.08	65					
Small chances	.16	.09	63					
Small likelihood	.18	.11	53					
Doubtful	.18	.12	47	.17	.15	.23	.21	.49
A small hope	.20	.15	44					
Somewhat doubtful	.34	.24	39					
Possibly	.35	.12	51	.49	.14	.32	.19	.74
Possible	.38	.12	50	.52	.18	.55	.23	.52
Perhaps	.39	.14	55	.45	.17	.43	.16	.18
A certain hope	.39	.20	31					
Not certain	.43	.22	36					
Uncertain	.44	.15	59					
Chance for	—	—	—	.50	.18	.44	.19	.30
Some doubt	.53	.24	32					
Danger (risk)	.53	.22	40	.57	.22	.57	.23	.73
Good chance	.65	.22	37	.75	.10	.61	.20	.56
Good hope	.66	.13	62	.77	.11	.66	.15	.53
Likely	.67	.16	40	.71	.15	.59	.18	.17
Assumedly	—	—	—	.75	.11	.64	.17	.46
Probable	.74	.13	54	.80	.13	.76	.11	.26
Most possibly	.75	.15	43					
Great chances	.76	.08	69					
Quite certain	.82	.11	50					
Small doubt	.83	.14	68	.90	.07	.84	.19	.01
Very good hope	.84	.10	57					
Very probable	.86	.08	70					
Certain	.92	.09	72					
Mean	.42	.12	56	.54	.12	.49	.17	.43

Results of Brun and Teigen's three-part testing numerical estimates of expression of uncertainty and perceived ambiguity [bru:88, p. 393]

Uncertainty *about* the content:
Linguistic markers of uncertainty

Expression	Group 1		Group 2		All subjects	
	Co-ord.	Prob.	Co-ord.	Prob.	Co-ord.	Prob.
Certain	1.1950	1.00	1.2952	1.00	1.2738	1.00
Possible	1.0897	0.96	0.8284	0.84	0.9105	0.86
Probable	0.8409	0.87	0.9252	0.87	0.9043	0.86
Expected	0.7239	0.82	0.7211	0.80	0.7133	0.79
Undecided	−0.5972	0.32	−0.3730	0.41	−0.4394	0.38
Uncertain	−0.7210	0.28	−0.8139	0.26	−0.7939	0.25
Improbable	−1.0741	0.14	−1.0435	0.17	−1.0610	0.16
Impossible	−1.4572	0.00	−1.5394	0.00	−1.5075	0.00

Co-ordinates and calculated probability points for the eight expressions of group 1, medical students (n = 26), group 2, other students (n = 52) and all subjects together (n = 78) [renooij:99,

	Expression	Probability (%)
I	Certain	100
II	Probable	85
III	Expected	75
IV	Fifty–fifty	50
V	Uncertain	25
VI	Improbable	15
VII	Impossible	0

Final scale with seven categories of probability expressions
plus their calculated probability points [renooij:99, p. 24]

Uncertainty *about* the content:
Linguistic markers of uncertainty

RATINGS OF EXPRESSIONS OF CONFIDENCE AND DOUBT, OVERALL AND FOR PAST

AND PRESENT PHRASINGS

Expression	Past		Present			
	M	SD	M	SD	t	d
I'm not sure, it's kind of...	2.88	1.08	2.91	1.24	-0.104	0.03
Oh, I don't know, I suppose it's...	2.94	1.07	3.02	1.25	-0.339	0.07
I suppose it could be...	3.13	1.12	3.34	0.96	-2.149*	0.44
I'm guessing, but I would say it's...	2.92	1.06	3.39	1.04	-0.958	0.20
I think it's.... isn't it?	3.08	1.31	3.48	1.41	-1.442	0.29
I think, I think it's....	3.19	1.03	3.61	1.37	-1.719	0.35
I could be wrong, but I think it's...	3.67	1.32	3.68	1.38	-1.411	0.29
I guess it's...	3.56	1.13	3.75	1.14	-1.602	0.33
It's.... I think.	3.40	1.18	3.75	1.22	-0.827	0.17
I'm not sure, but it may be...	3.44	1.27	3.84	1.14	-0.032	0.01
I'm not certain, but it could be...	3.65	1.12	4.14	1.07	-2.668**	0.53
I think it's.... but I can't be sure.	3.48	1.16	4.16	1.33	-2.149*	0.44
I can't say for sure, but I think it's...	3.81	0.99	4.16	0.94	-1.773	0.36
I'm not completely confident, but I think it's	4.12	1.02	4.20	0.90	-0.449	0.08
I think it's...	4.06	1.06	4.66	0.99	-2.865**	0.57
I could be mistaken but I'm sure it's...	4.17	1.23	4.68	1.07	-2.137*	0.43
I suspect it's...	4.25	1.12	4.68	0.98	-1.992*	0.40
I would say it's...	4.29	1.05	4.70	1.02	-1.952*	0.39
I believe it's...	4.56	1.11	4.86	1.11	-1.344	0.27
I remember it's...	5.25	1.14	5.18	1.08	0.299	-0.06
I'm fairly confident it's...	5.25	1.10	5.32	0.80	-0.342	0.07
I have no doubt, I mean I'm sure it's...	5.40	1.71	5.95	1.03	-1.869	0.38
I'm sure it's...	5.52	1.16	6.02	1.02	-2.232*	0.45
I have no doubt it's...	5.88	1.32	6.30	1.05	-1.665	0.34
I'm confident that it's...	6.06	1.18	6.43	0.97	-2.420*	0.49
I know it's...	6.08	1.49	6.45	1.00	-1.676	0.34
I know for a fact that it's...	6.25	1.37	6.50	1.00	-2.339*	0.47
I'm certain it's...	5.90	1.61	6.55	0.76	-1.428	0.28
I'm positive it's...	5.96	1.53	6.57	0.85	-1.005	0.21
I'm absolutely certain it's...	6.33	1.45	6.61	0.97	-1.116	0.22

Note.— The cues listed here are in the present tense. d is Cohen's d measure of effect size. df = 94.
* $p < .05$, ** $p < .01$

Wesson and Pulford's weighting with focus on the effects of time (present, past) on listeners' rating of expressions of confidence and doubt [Wesson:09, p. 155

Uncertainty *about* the content:
Linguistic markers of uncertainty

Boosters and downtoners:

Assigning a numerical weight, say, on a scale from 0 (impossible) to 100 (fact) would result in weights in which this relation is true:

somewhat likely < likely < very likely

Of course, there is the **reverse** effect when we use *somewhat* and *very* with the modal adverb *unlikely*:

very unlikely < unlikely < somewhat unlikely

Negation is, as usual, a bit more complicated. In d) below, we see the ordering of *unlikely* and *likely* and their negations.

unlikely <= not likely < not unlikely < likely

Uncertainty *about* the content:
Linguistic markers of uncertainty

"Different manners of acquiring knowledge correspond to different degrees of certainty about the truth of the proposition"

Willett[1988] proposed the following ranking in his study of various languages which have grammaticalized forms of evidentiality:

personal experience ≫ direct (sensory) evidence ≫ indirect evidence ≫ hearsay

DeHaan [deHaan2001] proposed a cross-linguistic comparison of source evidentiality:

sensory > inferential > quotative

Uncertainty *about* the content:
Linguistic markers of uncertainty

It was clear there is a more or less universal ranking for all of the lexical elements – even if there is no universal weight.

The major problem was finding a way to weight the various lexical elements, including boosters, downtoners and other evidential elements in a way that made sense – i.e., that the effects of boosters and downtoners would result in the correct ranking.

Uncertainty *about* the content:
Linguistic markers of uncertainty

The answer came via Holmes [1982]

Certain I	Probable II	Possible III

Maximum certainty that p is true or not true ← → Minimum certainty that p is true or not true
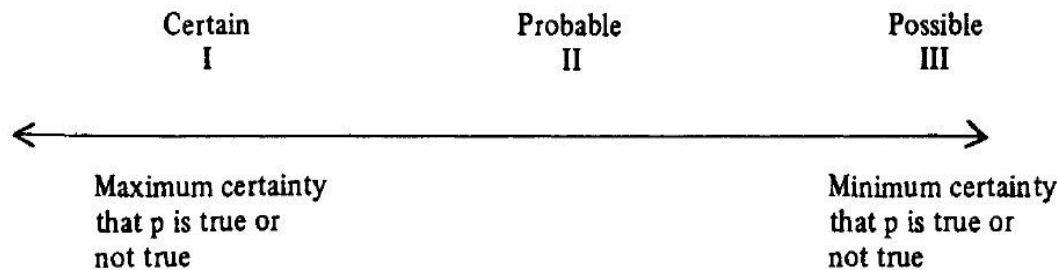
The following categories provide a relatively simple yet useful means of describing degrees of certainty expressed in English:

I. Certain: speaker asserts with certainty that the proposition is true or not true.

II. Probable: speaker asserts that the proposition is probably true or not true (i.e. improbable). .

III. Possible: speaker asserts that the proposition is possibly true or possibly not true.

Uncertainty *about* the content:
Linguistic markers of uncertainty

| p is *untrue* | I Certain p is *untrue* | II Probable p is *untrue* | III Possible p is *true* p is *untrue* | II Probable p is *true* | I Certain p is *true* | p is *true* |
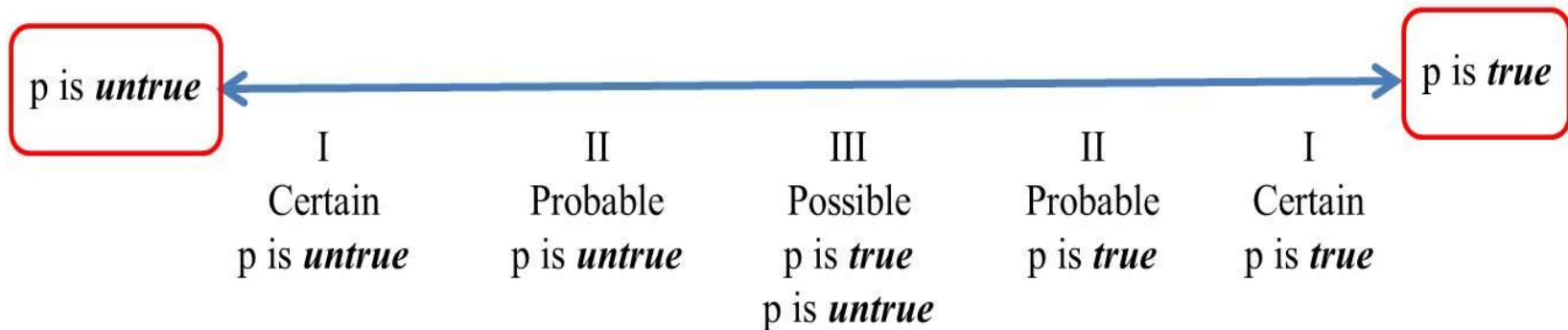
Holmes' scale opened so to reflect a *p is true* is at one end of the scale and *p is untrue* is at the other

Uncertainty *about* the content:
Linguistic markers of uncertainty
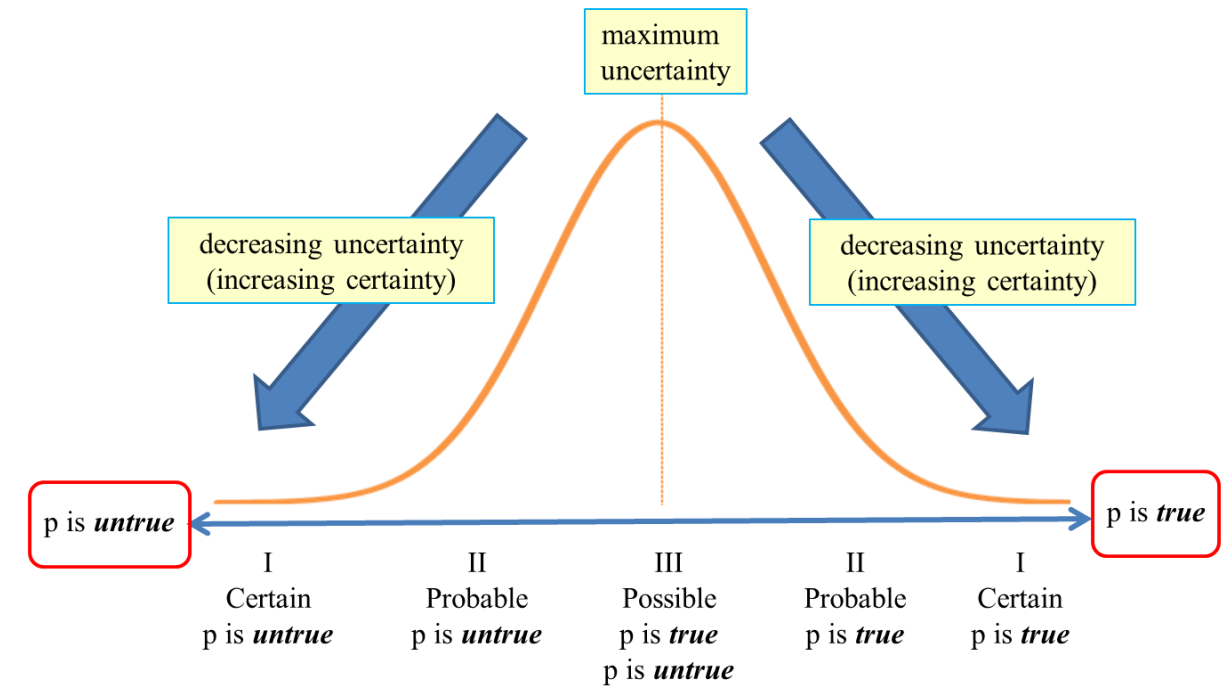
Uncertainty *about* the content:
Linguistic markers of uncertainty
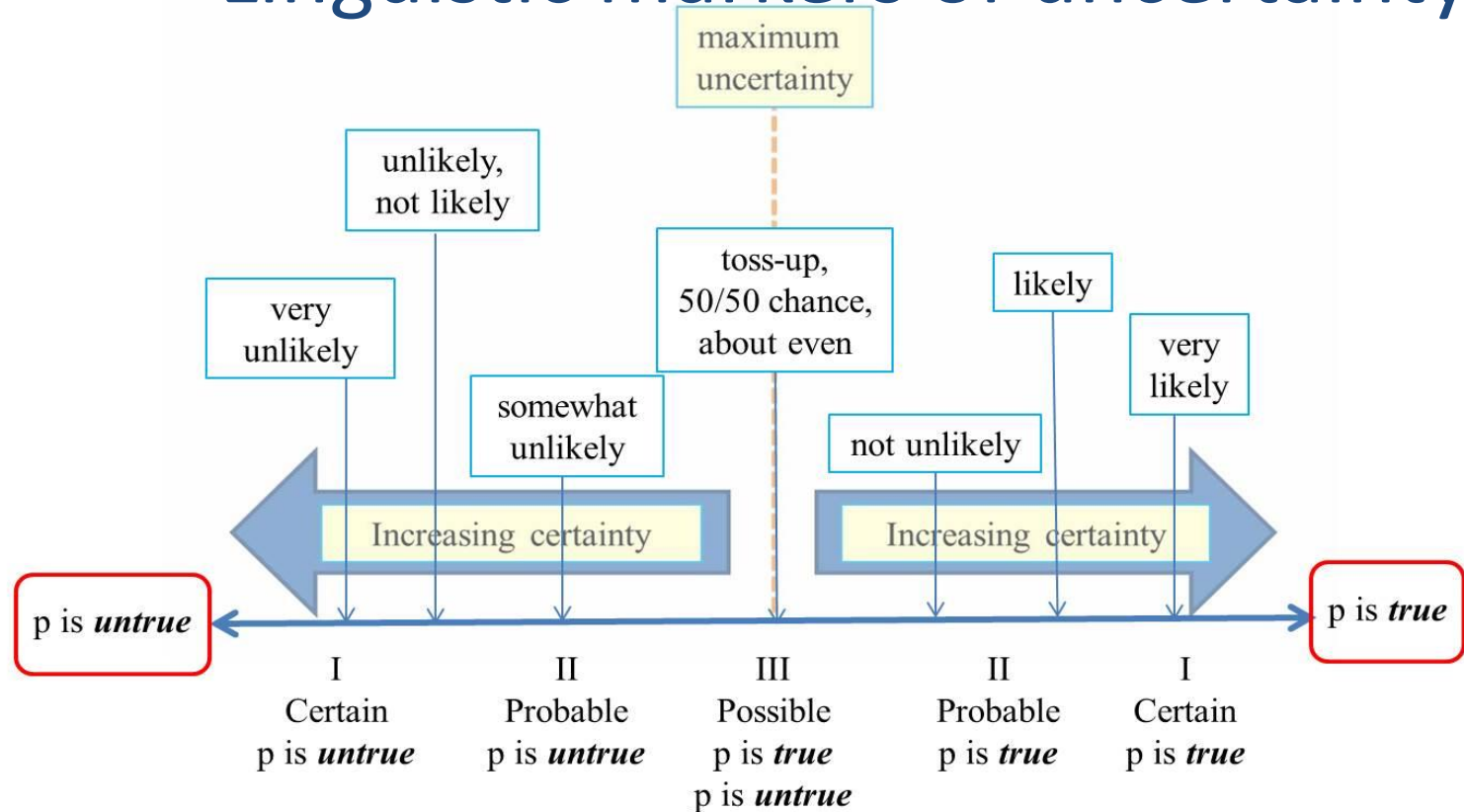
Overlaying some sample hedges onto the annotated scale

Uncertainty *about* the content:
Linguistic markers of uncertainty
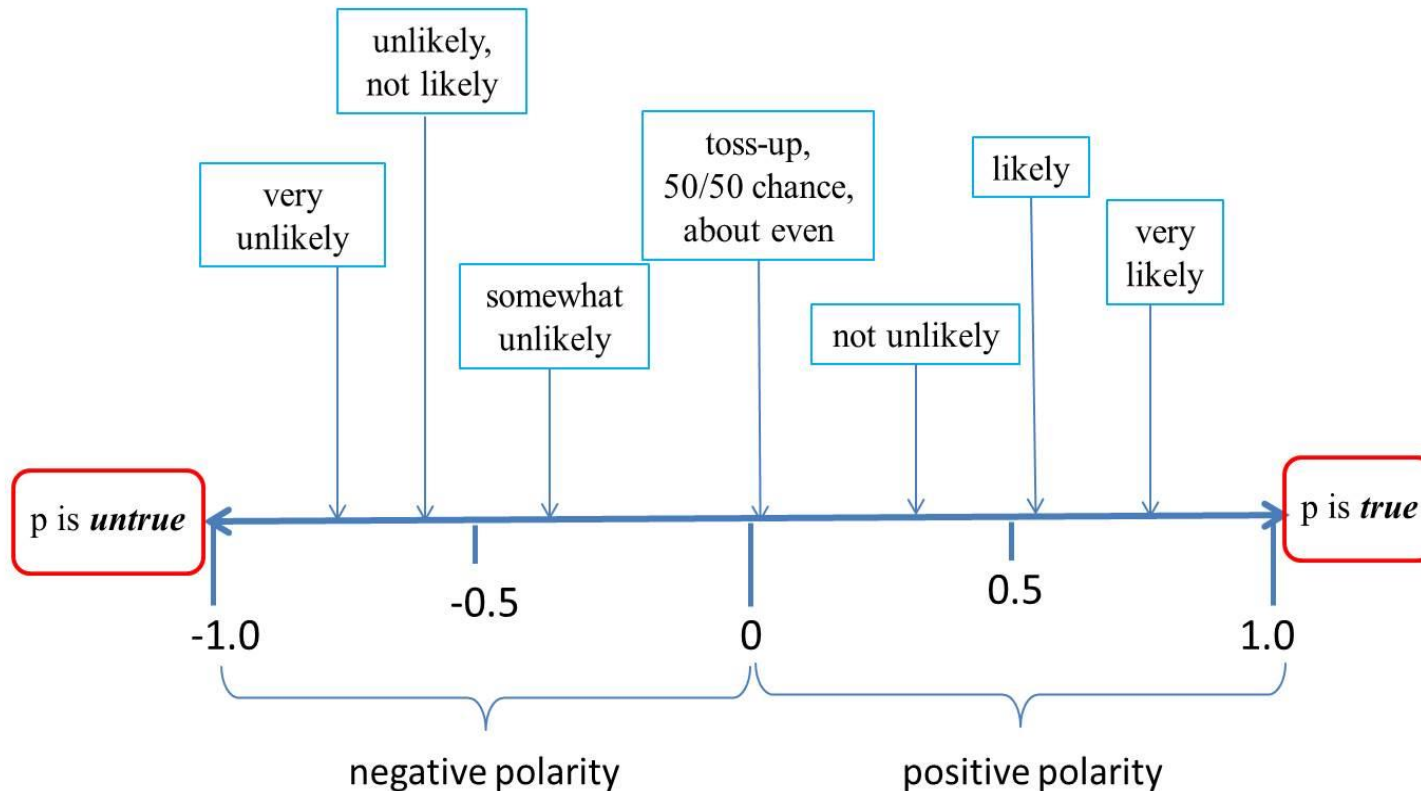
Using the point of maximum uncertainty, elements to the right are said to have positive polarity, whereas elements to the left have negative polarity.

Putting it all together

unlikely,

The goal is not to achieve numerical values but rather to achieve a relative ranking of various simple and compound evidential expressions

p   *true*

-1.0   0   1.0

negative polarity   positive polarity
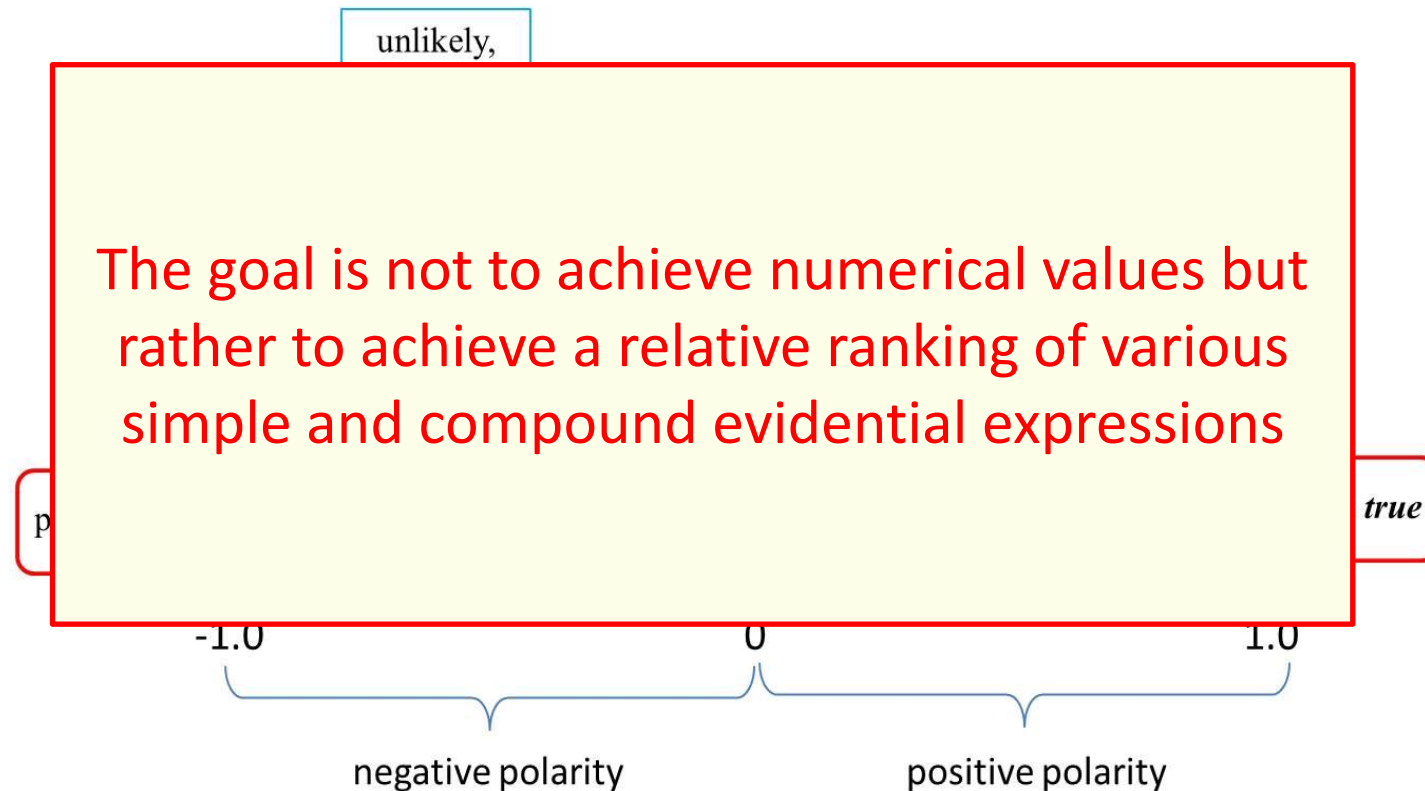
Putting it all together

Since we have already established that there is no "universal value" for a hedge (unless, of course, it is specifically stated as in "a 75% likelihood"), we will assign weights to some of the (unmodified) hedges as follows:

$$w_{likely} = 0.6$$

$$w_{unlikely} = -0.6$$

Putting it all together

Similarly, there are no "universal values" for boosters and downtoners. However, both types of modifiers vary in the intensity by which they strengthen or weaken the underlying value of the hedge

"*Extremely*" produces a bigger booster effect than "*very*" and "*somewhat*" has a very weak effect. Thus, we can assign (arbitrary) weights to these modifiers to reflect the relative degree of modification.

Thus,

   "*extremely*" would be assigned a weight of 0.3,

while

   "*very*" is weight with 0.2 to reflect its relatively weaker effect,

and

   "*somewhat*" has a relatively minimal effect.

Putting it all together

We have set the generalized form for the effect of the modifier on the original hedge to:

$$effect_{\text{modifier}} = \boldsymbol{p}(\boldsymbol{m} * (1 - |\boldsymbol{w}|))$$

where

$\boldsymbol{p}$ is the polarity of the hedge in question,
$\boldsymbol{m}$ is the weight of the modifying booster or downtoner and
$\boldsymbol{w}$ is the weight assigned to the hedge.

The effect of the term $(1 - |\boldsymbol{w}|)$ is to ensure that the resulting values of the modified hedges do not exceed the maximum limits (1.0 and -1.0) on the scale.

The use of the polarity $\boldsymbol{p}$ is to account for the differing behavior of the modification depending on the polarity of the hedge: for example, using a booster on a positively-poled hedge results in a value to the *right* of the original, whereas a booster on a negatively-poled hedge results in a value to the *left* of the original

Putting it all together

To demonstrate, we assign the value 0.3 to the booster "*very*" to represent the amount we believe the booster increases value of the hedge it is modifying.

When we multiply the weight w_{likely} by the *effect*$_{very}$ we end up with the following result:

$$w_{likely} + effect_{very} = 0.6 + (1)(0.3*(1 - |0.6|)) = 0.72$$

which indicates that "*very likely*" ends up to the right of "*likely*" as expected.

Similarly, when we modify the weight $w_{unlikely}$ by the *effect*$_{very}$ we end up with the following result:

$$w_{unlikely} + effect_{very} = -0.6 + (-1)(0.3 * (1 - |-0.6|)) = -0.72$$

which indicates that "*very unlikely*" ends up to the left of "*unlikely*" as expected.

Putting it all together

Likewise, to the downtoner "**somewhat**" we assign the value

$$effect_{somewhat} = -0.1,$$

which, we believe, reflects its weakening ("negative") effect on hedges.

When we multiply the weight w_{likely} by the $effect_{somewhat}$ we end up with the following result,

$$w_{likely} + effect_{somewhat} = 0.6 + (1)(-0.1*(1 - |0.6|)) = 0.56$$

and when we multiply the weight $w_{unlikely}$ by the $effect_{somewhat}$ we end up with the following result:

$$w_{likely} \quad effect_{somewhat} = -0.6 - (-1)(-0.1*(1 - |-0.6|)) = -0.56$$

which indicates that "**very unlikely**" ends up to the left of "**unlikely**" as expected.

Uncertainty *about* the content:
Linguistic markers of uncertainty
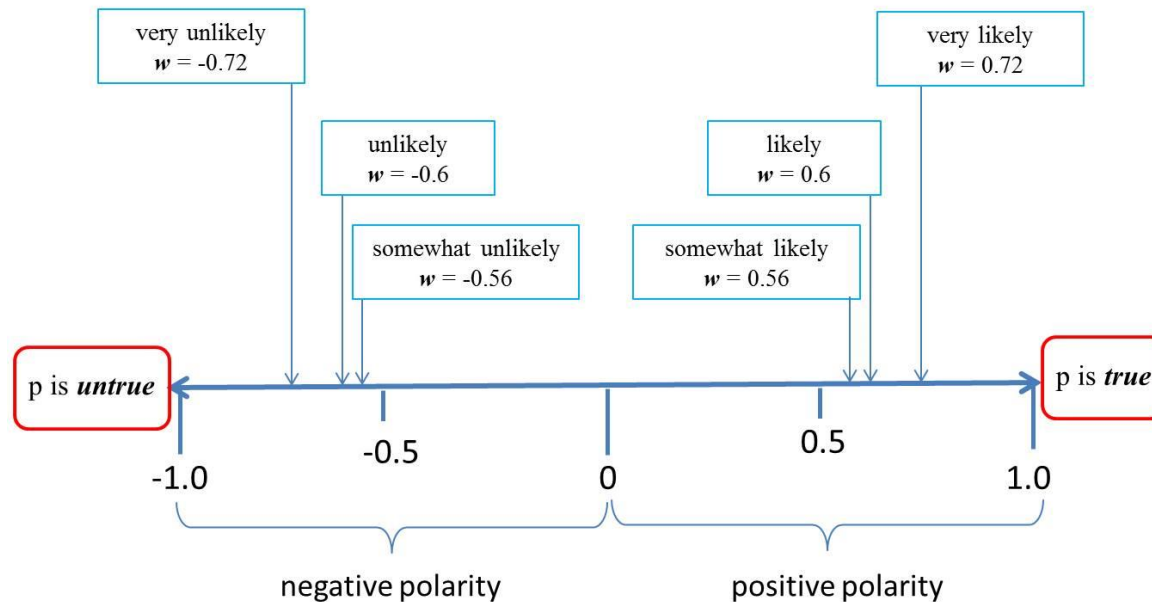
Putting it all together

Negation is generally straightforward: negating a hedge simply flips it around the (imaginary) axis of the point of highest uncertainty by changing its sign, that is, by multiplying by -1:

$$\textit{effect}_{\text{negation}} = -1$$

Thus, for "***not likely***" we end up with:

$$\textit{w}_{\text{likely}} * \textit{effect}_{\text{negation}} = 0.6 * -1 = -0.6$$

which is essentially the same as "*unlikely*."

Uncertainty *about* the content:
Linguistic markers of uncertainty
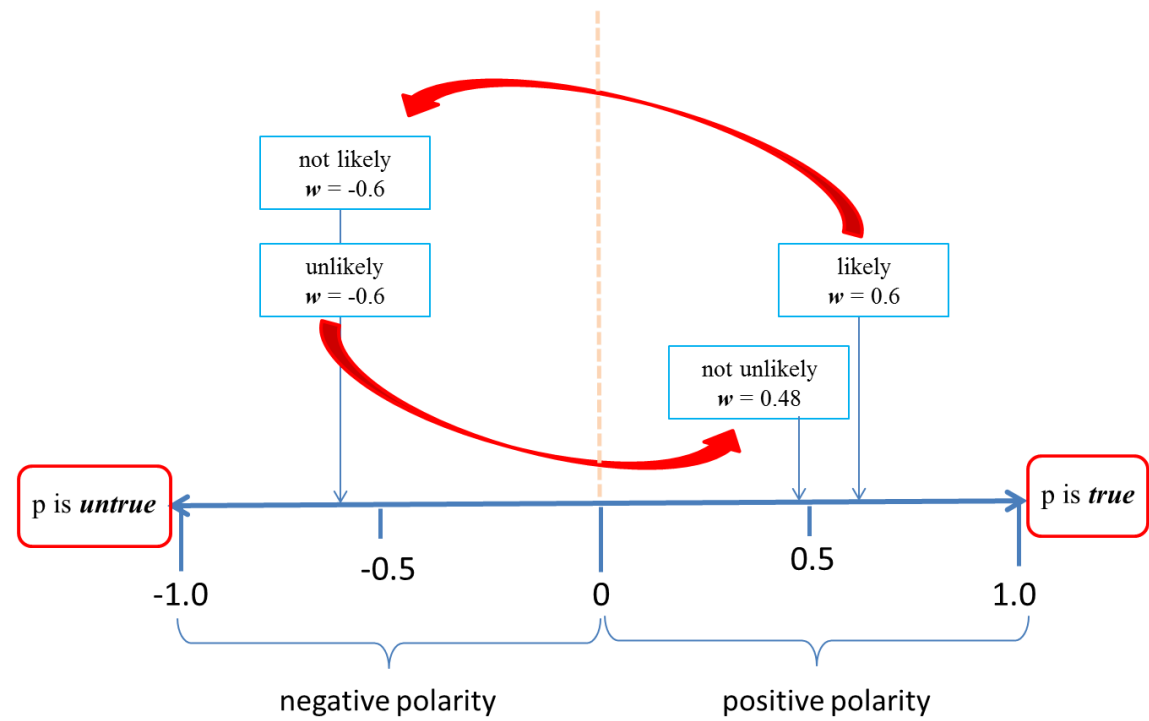
Putting it all together

In English, however, negation of a negatively-poled hedge generally ends up as somewhat softer than its opposite, in the same way that "*not ugly*" implies something less than "*beautiful*".

That is, "*not unlikely*" is usually considered weaker (and therefore should be closer to the point of highest uncertainty).

Thus we can differentiate the two:

If we assign the weight of the downtoner associated with negation of a negatively poled hedge to 0.2, the negation of "*unlikely*" results in

$$w_{unlikely} * effect_{negation} = -0.6 * (-1 + 0.2) = 0.48 ,$$

thus placing "*not unlikely*" to the left of "*likely*" and closer to the point of maximum uncertainty, indicating its relative weakness.

Uncertainty *about* the content:
Linguistic markers of uncertainty
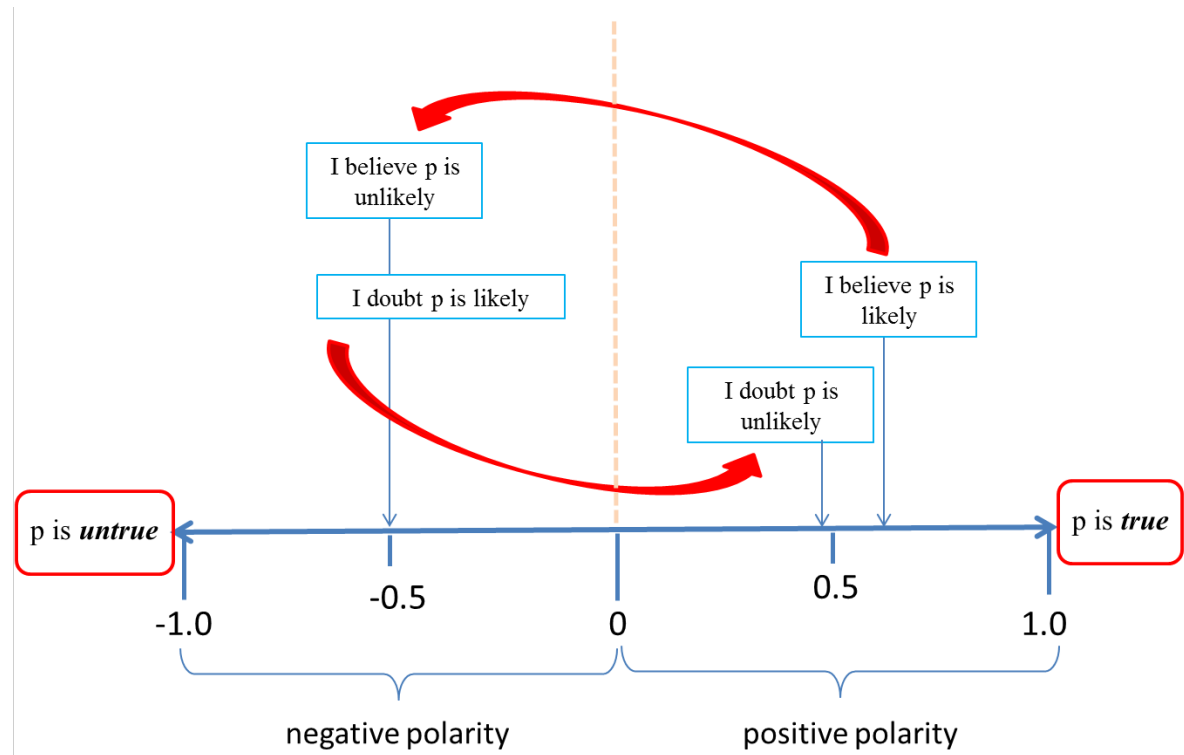
Point of maximum uncertainty

There is one place where boosters and downtoners do not function as described above: the point of maximum uncertainty. Mathematically speaking, this is not a surprise: since this point is assigned the value 0, and therefore all hedges associated with the point likewise are 0, then multiplication by any factor associated with any of these modifiers will cause it to remain 0.

There are some exceptions to this: for example "*better than 50/50*" implies that the associated value has moved from 0 into the positively-poled area. Similarly "*less than 50/50*" implies that the associated value has moved from 0 into the negatively-poled area. One solution may be to treat such expressions such expressions as compound hedges, rather than as normal hedges; for example "*better than 50/50*" could be assigned a value of 0.1, and "*less than 50/50*" assigned -0.1. An alternate solution would be to write separate rule for effects on elements located at the point of maximum uncertainty which would insert a polarity based upon whether the modifier was a booster or a downtoner

Uncertainty *about* the content:
Linguistic markers of uncertainty
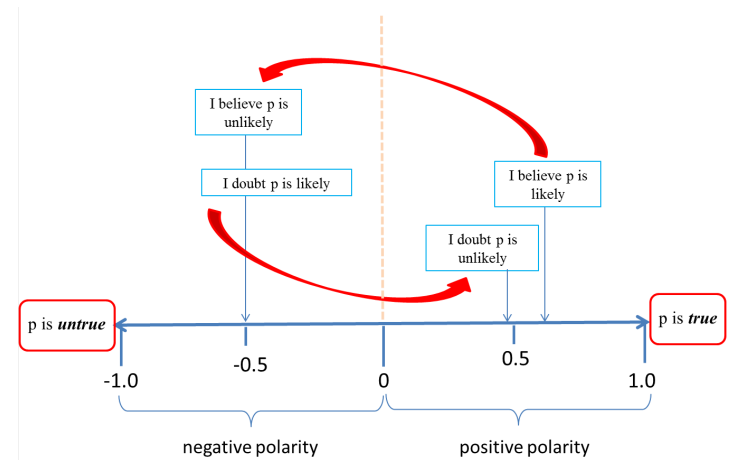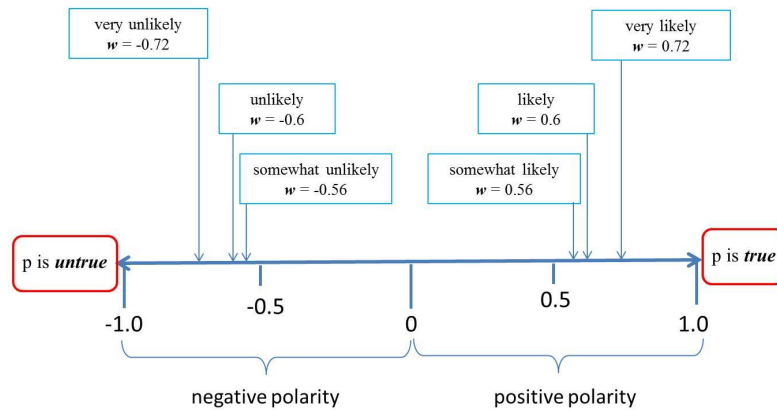
$$e = \prod_{i=1}^{n} w_{hedge_i} * \prod_{j=1}^{m} w_{hearsay/mindsay_j}$$

Uncertainty *about* the content:
Linguistic markers of uncertainty
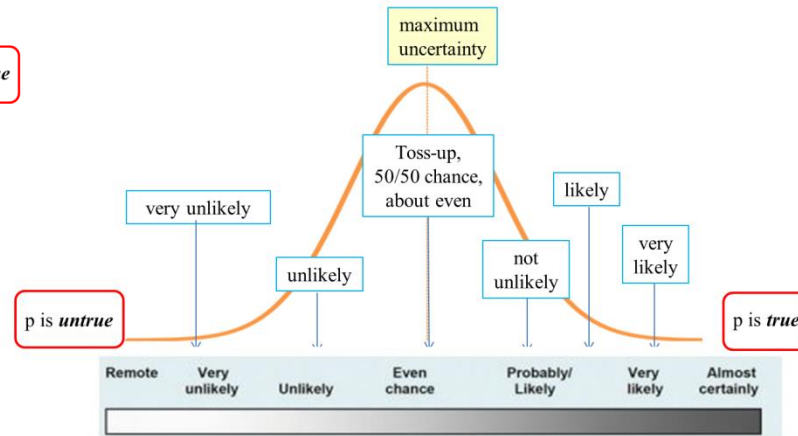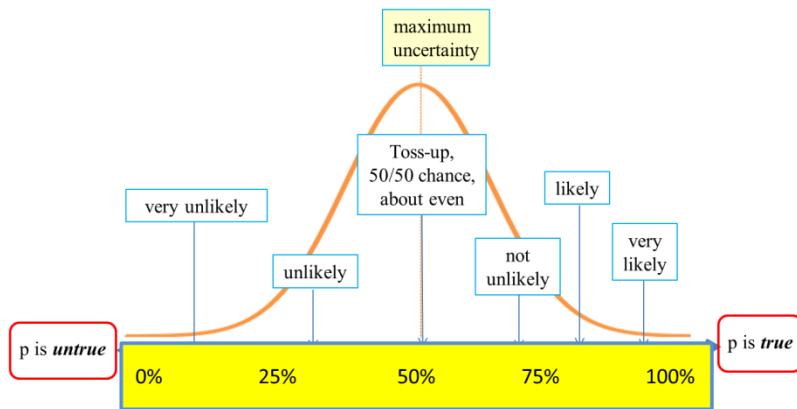
- **Questions ?**

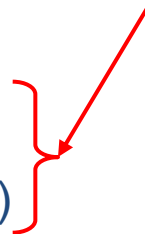# Uncertainty in the fusion process

There are several types of uncertainty :

1. Source uncertainty (how reliable is the source?)

2. Content uncertainty (how reliable is the content?)

3. Correlation uncertainty (how certain is it that various reports are related?)

4. Evidential uncertainty (how strongly is our information indicative of a specific threat?)

5. Model uncertainty (even with all factors present, how certain are we that the model mirrors reality?)

Uncertainty in the fusion process

There are several types of uncertainty :   data level

1. Source uncertainty (how reliable is the source?)

2. Content uncertainty (how reliable is the content?)

3. Correlation uncertainty (how certain is it that various reports are related?)

4. Evidential uncertainty (how strongly is our information indicative of a specific threat?)

5. Model uncertainty (even with all factors present, how certain are we that the model mirrors reality?)

Uncertainty in the fusion process

There are several types of uncertainty :

data level

1. Source uncertainty (how reliable is the source?)

2. Content uncertainty (how reliable is the content?)

fusion level

3. Correlation uncertainty (how certain is it that various reports are related?)

4. Evidential uncertainty (how strongly is our information indicative of a specific threat?)

5. Model uncertainty (even with all factors present, how certain are we that the model mirrors reality?)

Uncertainty in the fusion process

There are several types of uncertainty :

data level

1. Source uncertainty (how reliable is the source?)

2. Content uncertainty (how reliable is the content?)

fusion level

3. Correlation uncertainty (how certain is it that various reports are related?)

4. Evidential uncertainty (how strongly is our information indicative of a specific threat?)

5. Model uncertainty (even with all factors present, how certain are we that the model mirrors reality?)

model level

Source uncertainty

"On the Internet, nobody knows you're a dog."