I believe it's possible it might be so....

Exploiting Lexical Clues for the Automatic Generation of Evidentiality Weights for Information Extracted from English Text

Inaugural-Dissertation
zur
Erlangung der Doktorwürde
der
Philosophischen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität
zu Bonn

vorgelegt von

Kellyn Kay Elizabeth Rein (geb. Kruger)

aus

Lansing, Michigan USA

Bonn, 2016

Gedruckt mit der Genehmigung der Philosophischen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

Zusammensetzung der Prüfungskommission:
Herr Prof. Dr. Thomas Klein (emeritus)
(Vorsitzender)
Frau Prof. Dr. Claudia Wich-Reif
(Betreuerin und Gutachterin)
Herr Prof. Dr. Ulrich Schade
(Gutachter)
Frau Prof. Dr. Barbara Schlücker
(weiteres prüfungsberechtigtes Mitglied)

Tag der mündlichen Prüfung: 3. Mai 2016

For my dad, who always believed I would
and
for my mom, who always expected me to.
Thanks.

Sir Humphrey: With Trident we could obliterate the whole of Eastern Europe.

Prime Minister: I don't want to obliterate the whole of Eastern Europe.

Sir Humphrey: It's a deterrent!

Prime Minister: It's a bluff. . . I probably wouldn't use it. . .

Sir Humphrey: Yes, but they don't know you probably wouldn't.

Prime Minister: They probably do.

Sir Humphrey: Yes, they probably know that you probably wouldn't but they can't certainly know!

Prime Minister: They probably certainly know that I probably wouldn't!

Sir Humphrey: Yes, but even though they probably certainly know that you probably wouldn't they don't certainly know that, although you probably wouldn't, there is no probability that you certainly would!

Prime Minister: What?

(BBC Series ***Yes, Prime Minister: The Grand Design***, first shown 9th January 1986 – with thanks to Hoye[2005] for the reminder)

Acknowledgements

With very great thanks to Prof. Dr. Ulrich Schade who supported me and kept me sane through this whole long, long process over many years. My very deep gratitude also goes to Prof. Dr. Claudia Wich-Reif for her cheerful flexibility and most kind support especially through the end phase of this thesis.

Honorable mention goes to my husband, Dr. Volker Rein, who was always there in the background, encouraging me and helping as best he could to keep me going, as well as to our son Robin who walked the dog so I could stay a bit longer and work a bit more.

Contents

List of Figures

Chapter 1

Introduction and Overview

"You know what people are like," my father said. "Someone says, 'I suppose Leonard Kitchens could have put the rifle in the gutter, he's always in and out of the hotel,' and the next person drops the 'I suppose' and repeats the rest as a fact."

Dick Francis, *The 10 lb. Penalty* [1997, p. 206]

1.1 Motivation

In the new world of "Big Data", the worldwide web, and electronic publishing, information is expanding at such an incredible rate that no one knows how to begin making inroads into all of the information available. According to Gunelius [retrieved July 12, 2014] based upon a graphic produced by the computer software company DOMO, this was where some of this new data was coming from in 2013:

Every minute:

- Facebook users share nearly 2.5 million pieces of content.
- Twitter users tweet nearly 300,000 times.

- Instagram users post nearly 220,000 new photos.
- YouTube users upload 72 hours of new video content.
- Apple users download nearly 50,000 apps.
- Email users send over 200 million messages.
- Amazon generates over $80,000 in online sales.

Putting into perspective what this volume means, Gunelius [retrieved July 12, 2014] further writes: "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." She continues the comparison: "all of the written works of mankind created since the beginning of recorded history in all languages equals 50 petabytes of information" remarking that, at the time of writing (2014) Google processed 20 petabytes of information per day as a comparison.

Today, nearly three years later, the creation of new data is continuing to escalate. It will be no surprise to the reader that no one knows how to make sense out of all this data and put it to appropriate use. An article from Business Insider claims that only about 0.5% of all data is currently analyzed:

> "The rate at which we're generating data is rapidly outpacing our ability to analyze it." Professor Patrick Wolfe, Executive Director of the University College of London's Big Data Institute, tells Business Insider. "The trick here is to turn these massive data streams from a liability into a strength." (Browning, 2015)

Professor Wolfe has hit the nail on the head: turning this data from a liability to a strength is an important goal.

Clearly, this volume of digital information is beyond the capability of the humans producing it to process it by hand. Because of the massively large amounts of new information being produced each year, we are becoming more and more reliant on using automated methods to help us try to make sense of this information by using the speed and power of computers to helping us analyze text to find nuggets of useful and pertinent information to increase our understanding of the world around us, and to locate patterns which help us make informed decisions or to identify developing threats in time to preserve lives, property and national security. Not all of this volume is text-based; photos, videos, music, podcasts, etc., make up a significant proportion.

Indeed there has already been much work in trying to make sense of it all: the fields of information extraction, information retrieval, document classification and sentiment analysis are just a few examples of such applications (for an excellent brief overview of a variety of such techniques, cf. Kao and Poteet [2007]). Some techniques work using structured data; for example, knowledge extraction techniques operating on structured information such as relational databases and ontologies can help to determine new information buried within. Symmetrical relationships are an example of this: if our database contains the information that John is married to Susan, we can derive the new (to us) information that Susan is married to John.

But what about all of the new information that is being churned out every day on myriad topics in online newspapers, journals and magazines, digitally published research papers, reports on government and corporate websites, blogs and ebooks? This information is structured into sentences, paragraphs and chapters of varying length, not into neatly labeled rows and columns stored in files of similar and related items. In order to use this information, we must first parse the words that are in each sentence and make use of patterns appearing in the text to begin making sense of the words.

There have been great advances in recent years in text analytics: technologies which analyze natural language text using linguistic, statistical and machine learning algorithms to automatically perform such tasks using certain characteristics of written human language to determine which documents may be pertinent to our investigations, identifying persons, places or organizations, and identifying fact about or relationships between persons, places and events.

However, in all of this, there is one area which still remains problematic: how certain can we be that the information we thus discover is "true." And how might we quantify the "truthfulness"?

To illustrate this point, let's consider the following sentences:

(1) *John is a terrorist.*

(2) *I believe John is a terrorist.*

(3) *Mary told me John is a terrorist.*

(4) *The CIA suspects that John is a terrorist.*

(5) *It simply isn't possible that John is a terrorist.*

(6) *Do you think John is a terrorist?*

(7) *If John is a terrorist, then I am the Queen of Sheba.*

Each of these sentences contains the same pattern ***John is a terrorist***. An algorithm trained to look for such patterns would result in the relation ***John IS-A terrorist*** as a result for each of them.

However, there is other information contained in nearly all of these sentences which gives us a reason to doubt the veracity of the "fact" of John being a terrorist. When (1) is processed: we have no reason, based on this sentence, *not* to consider this a fact. (NB: we may have reason to doubt the accuracy of the source of this sentence which would affect our ultimate decision, but this is another discussion.). All of the remaining sentences include elements which insert some uncertainty about the truth of John being a terrorist. In (2) the writer expresses his personal opinion about the situation. Both (3) and (4) express hearsay, as the writer is repeating information received from third parties. Additionally, in (4) the use of the verb "to suspect" indicates uncertainty on the part of the original source about the opinion being expressed. In (5) the writer expresses her conviction that John being a terrorist is untrue, while (6) is a question rather than a statement, and as such requests rather than conveys information. And lastly, (6) is a conditional, and therefore uncertain as a result; additionally, it may be noted, that the statement contains sarcasm, which is intended to convey the speaker's strong doubts that John could be a terrorist.

These examples illustrate how the "credibility" of the propositional content *John is a terrorist* relies on linguistic clues contained in each sentence. In this case, the credibility ranges from "fact" (1) to doubt (5) to disbelief (7) with varying shades in between, while (6) might be interpreted as complete lack of knowledge as to the truth of John's status.

It is clear from not only the examples above, but also from the quote with which this chapter begins, that, if we plan to make use of extracted information, it is not sufficient to simply focus on algorithms that identify patterns in text in order to extract facts and meaning from textual information. We cannot treat all information as equally valid: we need to be able to determine whether that which we extract is credible. Before storing this extracted information in databases or ontologies, for use in models or to fuse with other pieces of information to help us make decisions, or recognize growing threats, we need to examine the clues embedded within the surrounding text to evaluate the truth of the assertion in the sentence.

However, recognition of such clues is just one part of the problem. The second part is determining how to evaluate these clues, and how to assign them some sort of weighting which reflects their relative credibility. In particular, as the results of the algorithms text extraction are often used within mathematical models which are designed to fuse the information acquired with other information in order to gain knowledge, it is of interest to assign (numerical) values which may be used by mathematical algorithms.

However, it turns out that in evaluating such clues and assigning them weights is complicated by the fact that even the humans who use and interpret these clues are not in complete agreement, as demonstrated by the following anecdote by Sherman Kent of the CIA:

> "A few days after the estimate appeared, I was in information 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'? What kind of odds did you have in mind?" I told him that my person 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. Understandably troubled by this want of communication, I began asking my own colleagues to the Board of National Estimates what odds they had had in mind when they agreed to that wording. It was another jolt to find 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." Kent [1964]

This clearly illustrates, as we will see in more detail later (cf. Chapter 6, *Towards Quantifying Evidentiality in Natural Language*), that even those individuals with similar training, background and working domains will not necessarily interpret such lexical clues identically. While this may complicate our task, it turns out there is some consistency in the relative ordering of such clues that we can exploit for our purposes.

The reader may ask, why is assigning a value reflecting the credibility of the information which we extract from text interesting or useful? To what end? In the following section we will discuss in more depth one of the important applications in which the credibility of extracted textual information plays a significant role.

1.2 Why Numerical Values?

As we move further into the first half of the 21st century, the world appears to be getting more and more unstable with the growth of terrorism, increasingly radical groups of individuals focused on religious, political or military dominance. A major concern for nations is the safety and well-being of citizens, the stability of the underlying financial and political systems on which these nations are based. One, perhaps even the most important, weapon against terrorism is information. Good, reliable information may help authorities to ward off deadly attacks, identify and arrest individuals involved, and to dismantle terrorist organizations to weaken them. Decisions made upon questionable, unreliable information may prove fatal — at a minimum, decision-makers must be aware that information might be uncertain in order to factor that appropriately into the decision-making process. For proactive responses, information about future events (by its very nature non-factual and uncertain) may play a significant role in, for example, apprehending ringleaders, or preventing an attack.

Additionally, in recent years the world has experienced a number of significant natural disasters and man-made crisis: the attacks on the the World Trade Center in 2001, the Indian Ocean tsunami of 2004, the 2010 earthquake in Haiti, to name just a few. Reliable information from various sources such as police, fire, military, hospitals, etc., is essential during the aftermath of the crisis. Particularly with the advent of social media, rumor, hearsay and deliberate untruths (trolling) propagate wildly through social media channels, complicating and even endangering rescue efforts. The ability to identify original observations from second- or third-hand information is important, to speculative information as speculation and not fact with help crisis managers separate the wheat from the chaff.

But why are we interested in assigning *numerical values* to indicate the uncertainness of extracted information?

Making sense of large volumes of information is increasingly being accomplished by computerized algorithms which analyze the information, determining relationships between various pieces of information, and identifying connections and patterns within the information which single events or activities of interest to the decision-makers. These algorithms sit on mathematical models and systems which evaluate the patterns to determine the likelihood that they point to particular actions or events. The determination of these likelihood is based upon weights which have been assigned to various aspects of the model, for example, how indicative a particular relationship is of a future threatening event. All data which is input into the system is weighted as to how strongly we believe the data to be factual or reliable: this is to off-set the well-known "garbage in, garbage out" effect. The algorithms which use the input return results that are also weighted as to the likelihood of a given event occurring; this likelihood is calculated in part using the weights assigned to the input. Therefore, we want to have a way to assign a numerical weight for the credbility of the extracted text information to be used by the fusion algorithms. (N.B., Chapter 6 *Mathematical Theories of Uncertainty*) provides an overview of some of the more well-known underlying mathematical systems used by such computer algorithms.)

When text-based information is extracted using natural language processing tools, it is pulled out of its context as we have seen in the example sentences earlier in this chapter. Once extracted, there are two ways of assessing this information. The first is essentially consider all extracted information as equally reliable, with the result that the system treats both speculative information and confirmed, reliable information as equally "factual" (or "non-factual" as the case may be). The second method is for human reviewers to examine the extracted information and provide some sort of credbility weighting. However, this second method has several weaknesses. One weakness is that weight assigned becomes a function of a given reviewer's personal belief about the veracity of the information, that is to say, the reviewer will rate it according to whether the information seems valid according to the reviewer's perception of the world. A second weakness, as we have seen in Sherman Kent's anecdote above, the numerical value of the weights can vary significantly from reviewer to reviewer (i.e., weights are inconsistent). In either case, the extraction process has removed any telling clues about the source and reliability of the information which might influence the reviewer's assignment of a numerical credibility value.

The human reviewing process has another weakness: the sheer volume of information which is being generated, and the time which would be needed to perform a reasonable analysis. Thus, it behooves us to try to find alternative methodologiesto support the process of assigning credibility weights to natural language information .

One obvious method for accomplishing this is to evaluate lexical clues intentionally embedded by the writer to strengthen or weaken the proposition contained in the sentence and use these as a basis for generating a credibility factor. That is the focus of this thesis.

1.3 Structure of This Thesis

Chapters 2 and 3 focus on the concept of ***uncertainty***. Chapter 2 *What is Uncertainty?* is an overview of *uncertaint*y as an over-arching concept. What exactly constitutes *uncertainty*? What are the various facets and expressions of this concept? How is it defined? It appears to be a "glass half full or half empty" question. It turns out, there is consensus on neither what exactly uncertainty is, or nor how other related concepts such as vagueness or imprecision fit into the picture. For some authors, anything below absolute knowledge is uncertain, for others such as Rubin [2010], *certainty* can be subclassified into absolute, moderate and low, with *uncertainty* ranking as the lowest level. For some authors *uncertainty* is a subcategory of other concepts. Similarly, some see *vagueness* as a form of *imprecision*, whereas some find it to be the other way around. Is ambiguity a subset of vagueness or or the other way around, or are, perhaps, the two conceptually different? And where does *fuzziness* fit into the picture? What about consistency and completeness? We will examine viewpoints from a number of authors from different domains as to how they define and classify various aspects of uncertainty, and how these are related. It seems that there is no single perfect answer for all cases. Indeed, what we will see is that the understanding of *uncertainty* and related concepts may vary depending upon the domain in which the researcher works.

In Chapter 3 we focus more specifically concepts and aspects of uncertainty in natural language, in this case, English. Here again we begin by examining the viewpoints of a number of authors about the manifestations and forms of uncertainty, both implicit and explicit, in English text. We then look at how various researchers define and categorize these language elements according to their uses in English. We end with an overview of English lexical items and formulations which are associated with uncertainty and from this we derive our classification of uncertainty in natural language for this thesis. This definition differentiates between uncertainty within the propositional content (*some*, *few*, *many*) and uncertainty about the propositional content (*possibly*, *unlikely*, *might be*), the latter of which is our focus.

In Chapter 4 *Evidentiality, Epistemic Modality, or Epistemic Stance?* we examine terminology used by linguists to describe the aspects of the uncertainty in the working description of uncertainty representation used in this work, namely epistemic modality, evidentiality and epistemic stance. As in the preceding chapters, there is some disagreement among experts as to the use of these terms. Based upon the discussion, as well as the practical goal of application of the results of this thesis, we select our terminology for the remainder of the thesis: evidentiality.

The mathematics of uncertainty is the main focus of Chapter 5 *Mathematical Theories of Uncertainty*, in which we look briefly at a selection of mathematical representations and theories used for the description and calculation of uncertainty, while Chapter 6 *Towards Quantifying Evidentiality in Natural Language* examines previous work done by researchers to apply measures of quantification to natural language lexical constructs which convey uncertainty. The main focus of Chapter 6 is, of course, on quantifying uncertainty *about* the propositional content.

In Chapter 7 *Putting It All Together* we use the discussions and conclusions of the preceding chapters as the foundation for an algorithm which can be used to calculate an evidentiality weighting for the propositional content of a statement based upon lexical clues in the sentence. Additionally, we demonstrate how the results may be mapped onto other credibility weighting scales for use by existing applications based upon mathematical systems such as those described in Chapter 5 *Mathematical Representations of Uncertainty.*

Finally, in the final chapter of this thesis (Chapter 8) we recap with our conclusions, identify open questions, and suggest areas of future work.

Chapter 2

What is Uncertainty?

The only relevant thing is uncertainty - the extent of our own knowledge and ignorance.

Bruno de Finetti, [1974, p. xi]

There are some statements that you know to be true, others that you know to be false, but with the majority of statements you do not know whether they are true or false; we say that, for you, these statements are uncertain.

David Lindley [2006, p. 1]

. . . As we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns – the ones we don't know we don't know. **Donald Rumsfeld** [retrieved 23 Aug 2014]

2.1 Introduction

Before beginning any attempt to describe a model for evaluating and quantifying uncertainty, one must first examine what is meant by *uncertainty*. This is less trivial than one may assume: like most things in our world, it depends upon your perspective. Defining, describing and modeling *uncertainty* is nuanced and filtered by experience, knowledge, application, belief, and even, to a certain extent, the languages we speak.

In this chapter, we will examine numerous approaches to formalizing and understanding *uncertainty*. Some are described in words, others in pictures. Some models focus solely on the concept of *uncertainty* (as demonstrated through its use as the root of hierarchical trees), other models see *uncertainty* as subsumed within a larger concept, such as *ignorance*, *imprecision* or *imperfection*.

Some models come from the field of information theory, from artificial intelligence, or from decision theory and economics. Others come from the fields of social sciences and computer graphics and ethics. Each provides a different viewpoint and, while some overlap, others differ quite dramatically from the others.

This is not intended to be an exhaustive survey of the field, but rather an overview of various methodologies with an eye to those which are of interest within the context of this thesis.

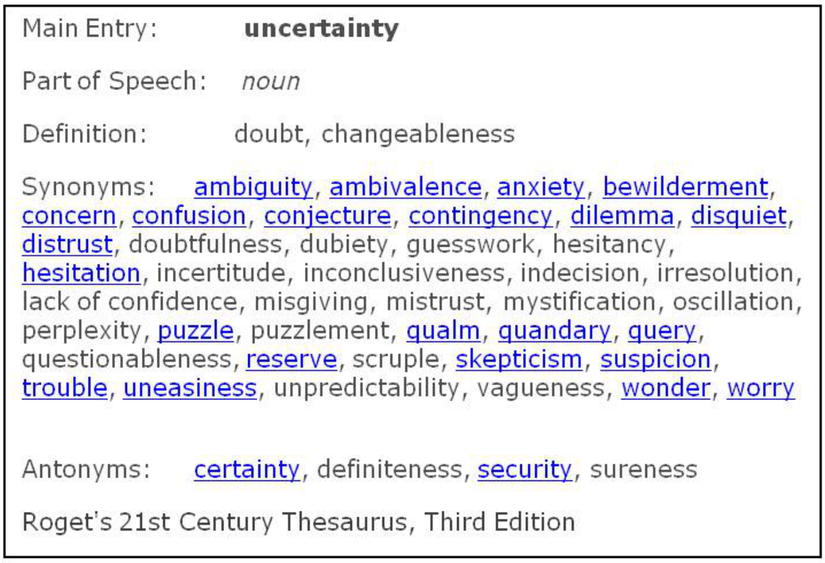
At the end of this chapter, in the interest of reducing misunderstanding and ambiguity (and thereby the reader's uncertainty) we will define many of the terms and concepts discussed in the following sections as they will be used further within this thesis.

2.2 Towards Defining Uncertainty

Uncertainty, in its popular, general sense, is defined by various English language dictionaries as

- *the state of being unsure of something* Webster's Online Dictionary [retrieved on Sep. 29, 2010]
- *the condition of being uncertain; doubt* The American Heritage Dictionary of the English Language, Fourth Edition [retrieved on Feb. 10, 2013]

- *the state of being uncertain* Oxford English Dictionary [retrieved on Sep. 29, 2010]


Main Entry: **uncertainty**

Part of Speech: *noun*

Definition: doubt, changeableness

Synonyms: ambiguity, ambivalence, anxiety, bewilderment, concern, confusion, conjecture, contingency, dilemma, disquiet, distrust, doubtfulness, dubiety, guesswork, hesitancy, hesitation, incertitude, inconclusiveness, indecision, irresolution, lack of confidence, misgiving, mistrust, mystification, oscillation, perplexity, puzzle, puzzlement, qualm, quandary, query, questionableness, reserve, scruple, skepticism, suspicion, trouble, uneasiness, unpredictability, vagueness, wonder, worry

Antonyms: certainty, definiteness, security, sureness

Roget's 21st Century Thesaurus, Third Edition

Figure 2.1: Entry for uncertainty in Roget's 21st Century Thesaurus, Third Edition [retrieved on Sep. 29, 2010]

Roget's Thesaurus (online) lists a myriad of synonyms for *uncertainty* as shown in Figure 2.1. From this, one can see at a quick glance, how widely diverse the numerous shadings of the word uncertainty are. Indeed W.J.M. Levelt [1989] and Clark [1987, 1988] maintain that there is no actual synonymy, as the core meanings of any two lexical items are not exactly identical (the uniqueness principle). Some of the synonyms are oriented more toward the psychological, that is, to a person's state of mind: *anxiety*, *bewilderment*, *confusion*, *uneasiness*, *worry*. Other synonyms reflect attitude: *doubt*, *mistrust*, *scepticism*, *ambivalence*, *suspicion*. This thesaurus entry makes one thing quite clear: there are many different ways to express *uncertainty* in words. Just as one selects the appropriate synonym from the thesaurus depending upon the shading required by the context in which the word will be used, the placing of the concept *uncertainty* within a conceptual model will depend upon domain in which it is being used, as well as each researcher's particular individual understanding of exactly what constitutes *uncertainty*.

When one considers that a number of entries appear as underlined hyperlinks in the synonym list, one may say that the entries displayed are in fact considered to be the most applicable synonyms for "uncertainty", that is, the entries displayed are in fact the most common synonyms for *uncertainty* in everyday usage. However, that there are other words and definitions with subtler or more domain-specific meanings that reflect a certain amount of uncertainty in their meanings that do not appear on this list. Within certain fields of endeavor such as statistics, economics, or information theory, there are definitions of and synonyms for uncertainty which are neither contained in the dictionary definitions above, nor in the thesaurus listing. An example of such specialized definition comes from the fields of economics: Knight, reflecting on the common – within that field – synonymic use of uncertainty and risk (the latter of which, it should be noted, does not appear in the Roget's listing above) states: "it will appear that a measurable uncertainty, or 'risk' proper, as we shall use the term, is so far different from an unmeasurable one that it is not in effect an uncertainty at all." F.H. Knight [1921] In other words, Knight differentiates between measurable and unmeasurable uncertainties, assigning the former the moniker *risk*." A number of elements are pertinent to the focus of this thesis — *ambiguity*, *lack of confidence*, *inconclusiveness*, *conjecture*, *vagueness* — which have to do with uncertainty in text are included in the list but other relevant synonyms such as *imprecision* or *nonspeicifity*, do not appear in the thesaurus listing.

Just as there are many different and specialized definitions of *uncertainty*, there are a multitude of ways of modeling and classifying uncertainty, each of which reflects the viewpoint of the researcher and his field of interest. For example, the typology of uncertainty of Smithson [1989], which we look at in more detail below, reflects his background in behavioral science, while the work of Klir and Wierman [1999] reflects issues in information theory and computational sciences. Both models use certain common terminology such as *ambiguity* and *nonspecificity*, but each has a different understanding and thus ordering of the concepts represented by these terms: Klir and Wierman [1999] view *nonspecificity* as a subset of *ambiguity*, while Smithson [1989] places *nonspecificity* under *vagueness*, rather than under *ambiguity*.

In the following subsections we will examine a number of typologies and models for classification of uncertainty which have been developed within various fields of research. Again, this survey is not intended to be exhaustive but rather to provide the reader with an overview of the variety of viewpoints conveying richness of diversity in thinking on this subject. These models have not, however, been selected simply at random: each of these models has been selected because each contains a point of relevance for the model being presented in this paper.

2.2.1 Uncertainty in Data (Information Theory)

Within the field of generalized Information Theory, Klir and Wierman [1999] have divided uncertainty into two subcategories: *fuzziness* and *ambiguity*, with the latter being further subdivided into the categories *strife* and *non-specificity* (Figure 2.2).

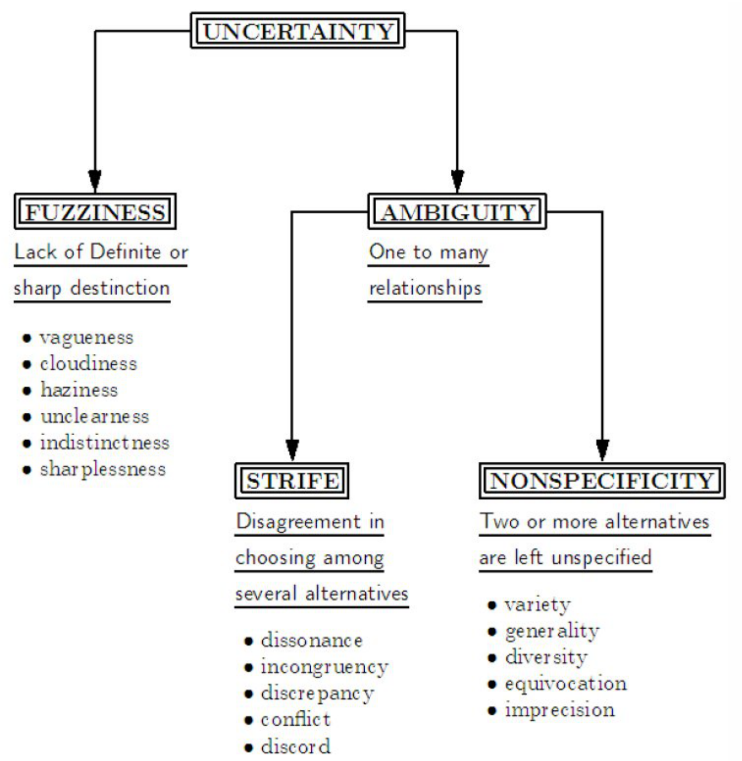


Figure 2.2: Typology of uncertainty [Klir and Wierman, 1999, p. 103]

The first level of distinction separates the less precise from the more precise. Fuzzy data are characterized by *lack of definite or sharp distinction*, with further synonyms including *vagueness*, *indistinctness* – in other words, there is a certain amount of imprecision in their values or descriptions.

The second subcategory, *ambiguity*, is defined as "one to many relationships" [Klir and Wierman, 1999, p. 103], which implies the problem of classification of the data under consideration rather than a lack of precision of value or description of an individual datum. In other words, data may have precise values, but their collective "meaning" may be imprecise. The subcategory of ambiguity which has been designated *strife* ("disagreement in choosing among several alternatives" [Klir and Wierman, 1999, p. 103]), describes situations in which data represent *dissonance*, *discrepancy* or *conflict*, i.e., that two or more data elements are to a certain degree contradictory. The second subcategory under ambiguity, *nonspecificity*, describes situations in which "two or more alternatives are left unspecified" [Klir and Wierman, 1999, p. 103], i.e., the data may be interpreted in more than one way, or point to more than one solution.

This model (Figure 2.2) concerns itself with data in systems, and appears to make the assumption that data elements are "true" even when they – their values or relationships between them – may be uncertain. There is no judgment as to the veracity of the content of the data. In other words, there is no provision in the model for classifying data which is false, deceptive, or otherwise partially or completely erroneous. For these very human aspects of an uncertain world, we need to look at other models.

2.2.2 Uncertainty Classification for Artificial Intelligence

Similar to the typology proposed by Klir and Wierman in the preceding section, Krause and Clark [1993] have focused on defining a classification for uncertainty which pertains to specific problems for the development of artificial intelligence applications (Figure 2.3). In contrast, however, Krause and Clark do not focus on uncertainty of data but rather on that which appears in the propositions manipulating this data.

This classification model first differentiates between types of uncertainty that apply to single propositions (*unary*) and those which apply to sets of propositions (*set theoretic*). Each of these is further subdivided into two major classifications, *ignorance* and *conflict*, which are then further decomposed into characteristics of each relevant to either single or multiple propositions.

While there are some similarities of these characteristics to those in Klir and Wierman (i.e., *ambiguity*, *strife/conflict*), there are some new considerations: the concepts of *anomaly/error*, *ignorance*, *incompleteness* and *irrelevance* which are missing from the previous model. One might argue that Klir and Wierman have tacitly included these concepts in their model under

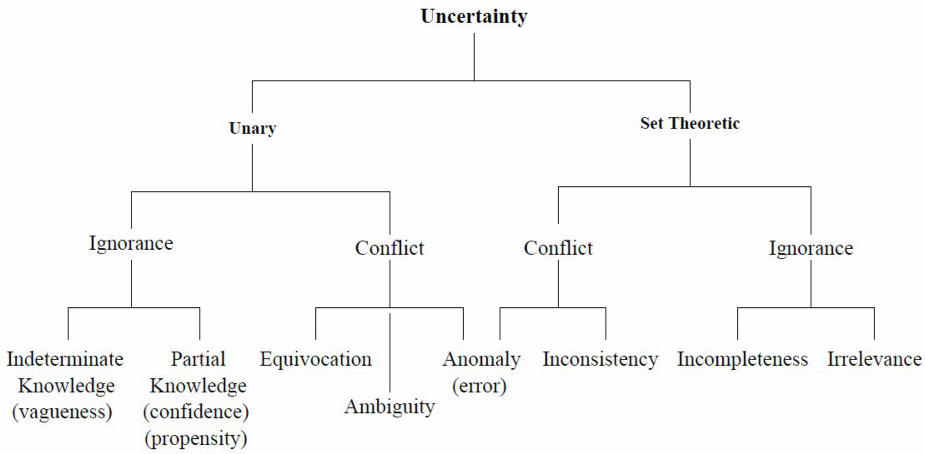


Figure 2.3: Uncertainty classification for artificial intelligence. [Krause and Clark, 1993, p. 7]

the classification *strife*. *Anomaly*, for example, could be viewed as *discord* – while *nonspecificity* could be seen as a side-effect of *ignorance* and *incompleteness*. However, such an interpretation may not have been the intention of the authors.

Certain elements contained within this model, such as the concepts of *ignorance*, *error*, and *irrelevance* are dealt with more explicitly by Smithson [1989] in the following section.

2.2.3 Uncertainty in knowledge (Social Sciences)

Smithson [1989] bases his *typology of ignorance* (containing elements of uncertainty) on his experiences in the area of cognitive science, which includes aspects of human behavior and unreliability (Figure 2.4). These are missing in the typology from Klir and Weirman, and only partially covered by Krause and Clark. They are, however, of paramount importance for the subject of this thesis.

Interestingly, the major change in viewpoint from the preceding models is the placement of *uncertainty* in this model. Whereas both of the preceding models focus on defining *uncertainty*, Smithson sees *uncertainty* in the middle of the larger concept of *ignorance*. All three models contain the term *ambiguity*; however, Klir and Wierman have designated this as one of their direct subcategories of *uncertainty*, as does Smithson, whereas Krause and

Clark see this as a subcategory of *conflict*.

There are further interesting differences between the various models. For example, Klir and Wierman designate *vagueness* as a subcategory of *fuzziness* whereas Smithson inverts the ordering. In the case of *fuzziness*, one might conjecture that this lies perhaps on a differing understanding of the term based upon expertise: in the one case *fuzziness* is very specifically defined as a mathematical model, the other definition is in a more generic sense of "lack of clarity."

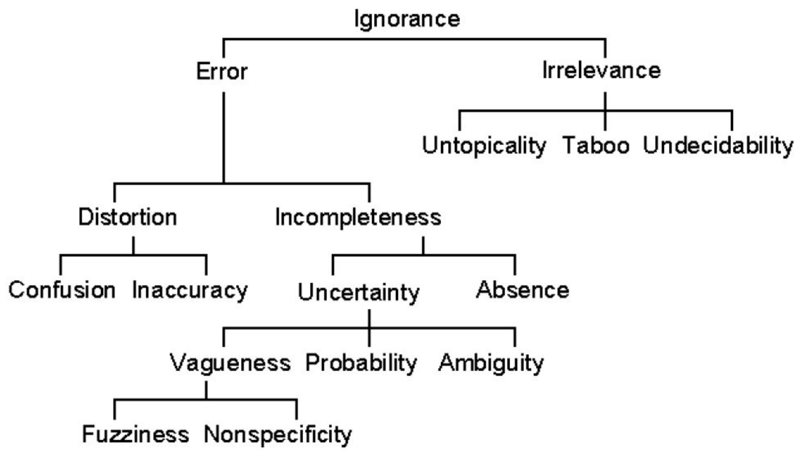


Figure 2.4: Typology of ignorance. [Smithson, 1989, p. 9]

Of interest in the Smithson model is that human factors play a significant role, in particular, in the sub-branch *irrelevance* (e.g., *untopicality* and *taboo*) and to a lesser extent in the subbranch *distortion*. When dealing with information derived from humans, as we will be doing within the scope of this thesis, one cannot ignore the vagaries, shortcomings and motivations of the individuals producing the information.

2.2.4 Imperfect information: Imprecision – Uncertainty

Smets is a leading name on the topic of uncertainty in artificial intelligence. He wrote many papers revolving around the overarching topic of *imperfection* in data or information. As shown in Figure 2.5, Smets [July 2, 1999] divides imperfection into three major areas: *imprecision*, *inconsistency* and *uncertainty*, further explaining:

"Imprecision and inconsistency are properties related to the content of the statement: either more than one world or no world is compatible with the available information. Uncertainty is a property that results from the lack of information about the world for deciding if the statement is true or false. Imprecision and inconsistency are essentially properties of the information itself whereas uncertainty is a property of the relation between the information and our knowledge about the world." [Smets, July 2, 1999, p. 2].

Imprecision he further breaks down into *imprecision without error* and *imprecision with error*. Ambiguous, approximate, fuzzy, incomplete or missing information is assumed to be "without error," that is, it is not incorrect, just unclear in some way. Imprecision with error he decomposes into *erroneous* or *incorrect* when it is completely wrong, *inaccurate* when the data is not correct in some way but the error is minimal, *distorted* when both inaccurate and invalid, *biased* when systematically distorted, and *nonsensical* or *meaningless* when dramatically erroneous.

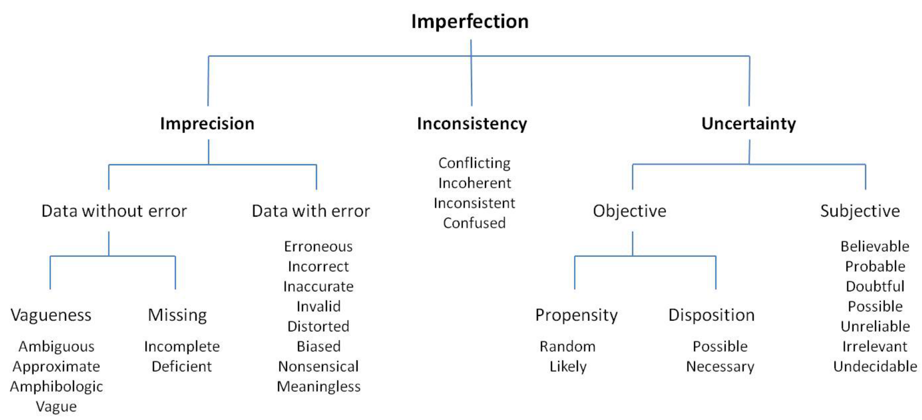


Figure 2.5: Adaptation of Smets typology of imperfect information by [Jousselme et al., 2003, p. 1211]

Furthermore, he divides uncertainty into *objective* and *subjective*. Objective uncertainty, he argues, is related to chance or randomness but he further refines it into *propensity* – how "likely" an event is to happen (measured as a probability), its *disposition*, the possibility of the event's happening. (NB: the mathematical concepts of probability and possibility will be discussed in a later chapter.) Subjective uncertainty is a measure of one's belief that this could occur, based upon one's understanding of the world.

Smets [July 2, 1999] provides a quite detailed "structured thesaurus of imperfection" which supports his model.

2.2.5 Epistemic Interpretations of Uncertainty

In Jousselme et al. [2003] the authors look at uncertainty for use in situation analysis from the perspective of military and security applications, and present an overview of several models for uncertainty and examine the applicability of those models to their field of interest.

Additionally, they add an overview of the quantification methods for the various interpretations of uncertainty, which provides an interesting insight into some of the variations of the models reviewed in this chapter. It is, therefore, interesting to take a brief look at this classification model, particularly as this is a generalization of the terms *objective* and *subjective* from Smets's model presented in the previous section.

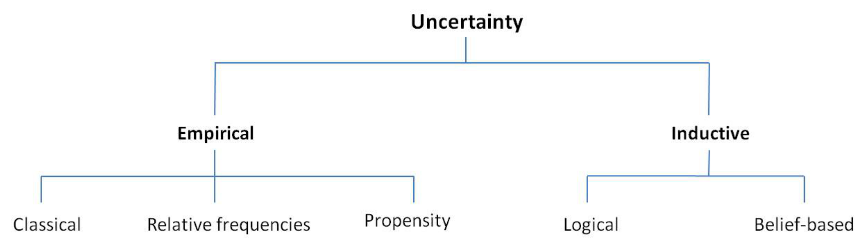


Figure 2.6: Epistemic interpretations of uncertainty. [Jousselme et al., 2003, p. 1212]

As shown in Figure 2.6, they divide the interpretations into *empirical* and *inductive*. *Empirical* implies experimentation and knowledge (internal representation) of possible states. Under *empirical* there are three categories:*classical*, *relative frequencies* and *propensity*. *Classical* implies complete knowledge of all possible outcomes and is based upon combinatorics and symmetry. *Relative frequencies*, on the other hand, assumes that complete knowledge is not available, but gives conditional results based upon the results of a large number of experiments. *Propensity*, which they use similarly to Smets, is the inherent preference or inclination toward a specific state.

The second type of uncertainty in this model is inductive, in which uncertainty is quantified, not through experimentation or knowledge, but by logic or by one's beliefs about possible outcomes (not necessarily based upon experience as in *propensity*).

Several of these concepts will be further discussed in later chapters in this thesis.

2.2.6 Imperfect Knowledge of the Information State

Gershon [1998] comes at uncertainty from yet a different angle: presentation of an imperfect world through visualization (graphics). The applied nature (visual presentation of information) of his model is obvious: certain aspects of his model appear synonymously in other models discussed above – inconsistency, incompleteness or inaccuracy – but they have been classified differently. Although the focus in his model is somewhat tangential to the topic of this thesis, there are a few aspects worth noting.

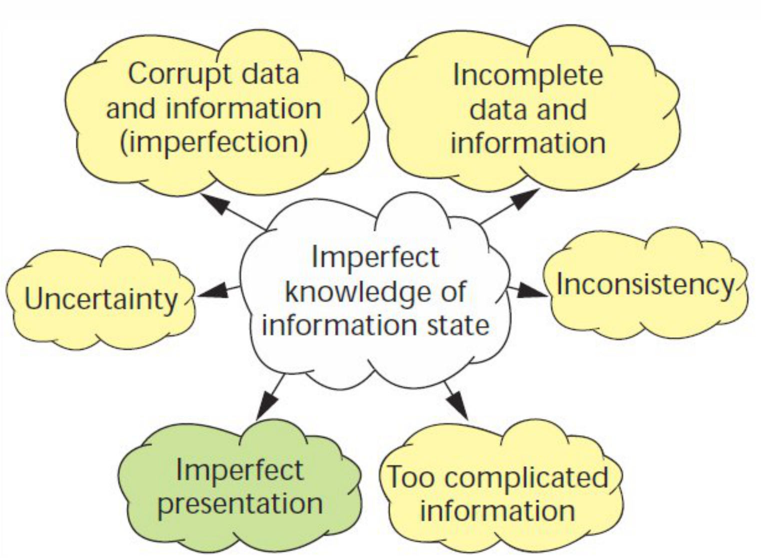


Figure 2.7: Taxonomy of the causes of imperfect knowledge of the information state [Gershon, 1998, p. 43]

For his purposes, Gershon [1998] defines *uncertainty* as "data and information [which are] known, but [where] the user is not sure about their existence or accuracy." This corresponds to elements of the model of Krause and Clark (partial knowledge, confidence). Gershon places *imperfect knowl-*

edge at the center of his model Figure 2.7. He views uncertainty as a result of imperfect knowledge, a view which correlates to a certain extent with Smithson.

Debatable in this model is the contention that imperfect knowledge of the information state results in corrupt data and information. In the sense that the term *corrupt data* is widely used to indicate data which is erroneous in some way, one could argue that this arrow in the diagram should be reversed, i.e., that corrupt data contributes to imperfect knowledge, rather than vice versa.

2.2.7 Typology for Information Uncertainty

In the intelligence area, the consequences of misjudging the accuracy of information or drawing erroneous conclusions may be, quite literally, fatal. Therefore, an understanding of sources and types of uncertainty in information is necessary. To this end, Thomson et al. [2005] have created a typology defining different categories of uncertainty pertinent to the fields of data analysis and intelligence.

Category	Definition
Accuracy/error	difference between observation and reality
Precision	exactness of measurement
Completeness	extent to which info is comprehensive
Consistency	extent to which info components agree
Lineage	conduit through which info passed
Currency/timing	temporal gaps between occurrence, info collection & use
Credibility	assessment of info source
Subjectivity	amount of interpretation or judgment included
Interrelatedness	source independence from other information

Figure 2.8: Categories in Analytic Uncertainty Typology. [Thomson et al., 2005, p. 152]

In contrast to the preceding models, which attempted to provide more abstract classifications of various aspects of uncertainty, Thomson et al. have quite pragmatically examined various sources of uncertainty. Figure 2.8 provides a listing of the categories identified, with a brief definition of each category, while Figure 2.9 decomposes each category into more detailed subcategories and provides a few examples of each category.

Additionally, Figure 2.9 provides a quite valuable checklist for investigating the various forms of uncertainty, particularly on the analytical level. As

Category	Subcategories	Examples
Accuracy/error	• Collection Accuracy • Processing errors • Deception	• Documents that are translated into English may contain translation errors. • A report may note that 50 tanks were observed although the tanks may in fact be dummy placements.
Precision	• Precision of collection capability	• resolution of satellite imagery
Completeness	• Composite completeness • Information completeness • Incomplete sequence	• images of a site may not be available on a particular day because of adverse weather conditions. • an intercepted conversation may have words that were not clear • the lack of confirming information might signal incompleteness
Consistency	• Conflict among info • Model/observation Consistency	• multiple sources or data types may actually conflict • models of events may differ from observations
Lineage	• Translation • Transformation • Interpretation	• Machine translation is more uncertain than human linguist translation • Measurements or signals may have been transformed • Information that comes directly from the collection capabilities has a different lineage than an interpretive report produced by an analyst
Currency/timing	• Temporal gaps • Versioning	• Images that show new objects do not show when the object first appeared • Time between when events occurred, when they were reported, and when the information is available to analysts • Reports may have multiple versions, sometimes with major changes.
Credibility	• Reliability • Proximity • Appropriateness • Motivation (of the source)	• Possibility of deliberate disinformation • Source may not have expertise on this subject • Information may be second hand
Subjectivity	• Analytic judgment	• Amount of interpretation added rather than pure facts
Interrelatedness	• Source independence	• Likelihood that the source derives from other reported information (such as repeated news stories)

Figure 2.9: Analytic Uncertainty Typology developed by [Thomson et al., 2005, p. 153]

a result, this list is quite closely aligned with the model for the evaluation of uncertainty in information fusion which we present in this thesis

2.2.8 Uncertainty in Decision-Making

Tannert et al. [2007] address uncertainty within the context of decision-making, in particular with a focus on the moral and ethical issues involved in decision-making based upon incomplete knowledge. Whereas Klir and Wierman[1999] focused on uncertainty in data and Smithson[1989 on the uncertainty in knowledge, Tannert et al. [2007] analyzed where uncertainty comes into the process of decision-making, with a strong focus on the ramifications of uncertainty upon decisions, in particular upon the consequences and risks resulting from these decisions. According to the authors, "each of the sub-forms [in their taxonomy] describes a particular type of mismatch between the knowledge required and the knowledge available for rational decision-making." [Tannert et al., 2007, p. 894] This mismatch is of particular interest within this paper, as the model presented is designed to support decision-making.

In their typology (Figure 2.10), Tannert et al. subdivided uncertainty into two major categories: *objective uncertainty* and *subjective uncertainty*. Objective uncertainty, based upon available knowledge and underlying, often complex, situations. In other words, objective uncertainty is based on "things that be", such as situational relationships and information which is not subjectively interpreted (or, in the case of "knowledge", is accepted as "truth"). Objective uncertainty is further decomposed into *epistemological uncertainty* ("gaps in knowledge") and *ontological uncertainty* ("caused by the stochastic features of a situation, which will usually involve complex technical, biological and/or social systems. . . often characterized by nonlinear behavior, which makes it impossible to resolve uncertainties by deterministic reasoning and/or research")[Tannert et al., 2007, p. 895].

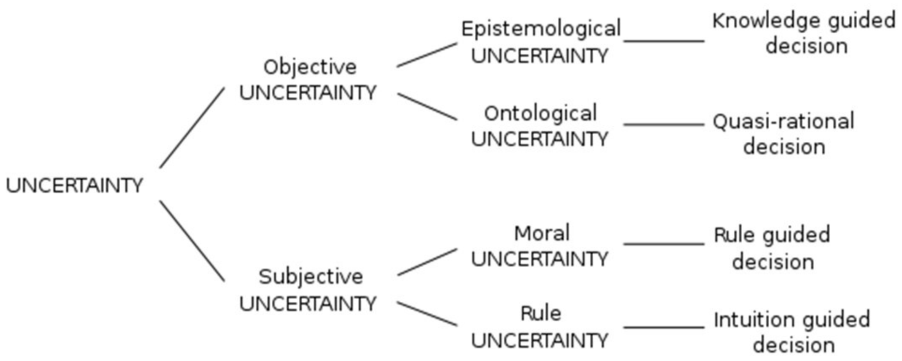


Figure 2.10: The taxonomy of uncertainties and decisions. [Tannert et al., 2007, p. 895]

Subjective uncertainty is "characterized by an inability to apply appropriate moral rules." [Tannert et al., 2007, p. 895] Within this classification the authors differentiate between two subcategories: *moral uncertainty*, in which there is a "lack of applicable moral rules" on which to base decisions, requiring decision-making by extrapolation from more generalized rules (i.e., interpretation); and *rule uncertainty*, in which decisions are made not based upon rules, but upon "internalized experiences and moral values." In short, subjective uncertainty results from the lack of clear, commonly accepted rules, and/or the need (or desire) to base decisions upon internalized values.

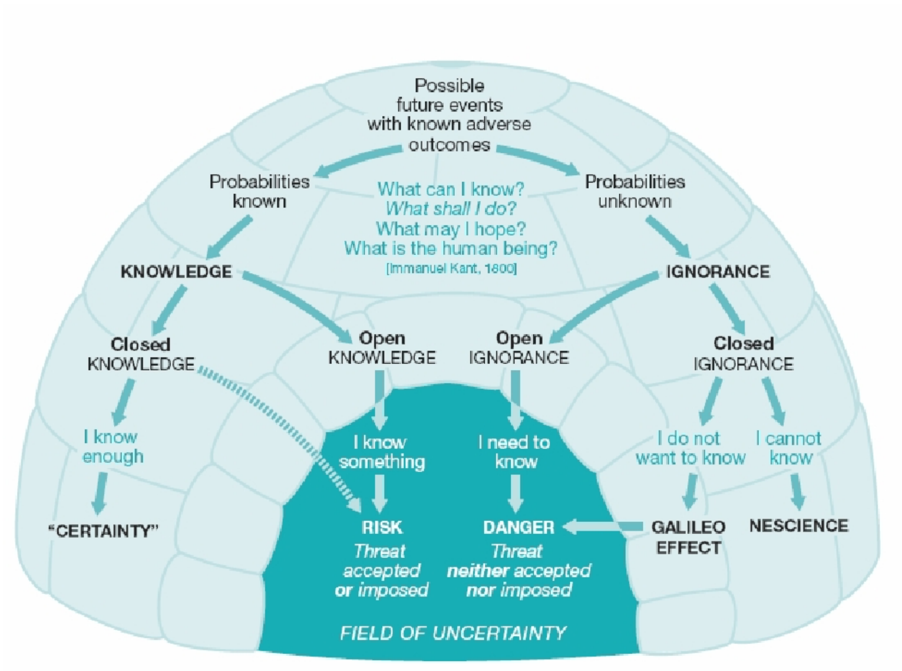


Figure 2.11: Igloo of uncertainty. [Tannert et al., 2007, p. 894]

In addition to their classification of types of uncertainty, Tannert et al. have created a schematic approach to represent the effects of this uncertainty upon decision-making as shown in Figure 2.11. In this representation, they return to some of the ideas of Smithson, in particular, *ignorance* and *indecidability* albeit with some modifications.

> "Our schematic approach, the 'igloo of uncertainty' ... mainly distinguishes between open and closed forms of both ignorance and knowledge. Within that framework, dangers are defined in terms of the possible outcomes of a given situation. To understand the potential adverse effects of a decision, we therefore require an approximation of the quality of dangers in any given event. Consequently, a rational approach is to give an estimate of the probability that the respective event will happen, and to

> assess the hazard and the possible impact of the event. Classical risk assessment then takes the product of probability and the expected hazard dimension to obtain a quantitative measure of risk." [Tannert et al., 2007, p. 894]

Interestingly, several elements of the Smithson typology appear in this igloo schema, albeit some in forms which are not immediately apparent: e.g., Smithson's *ignorance* appears as Tannert's undecidability *("I cannot know")*. Here it should also be noted that the igloo model, while conceived with ethical considerations in mind, also illustrates the connection between risk and uncertainty as classically defined within the field of economics.

2.2.9 Uncertainty in Linguistic Data

In Auger and Roy [2008] the authors discuss uncertainty within the framework of linguistic data. They examine a number of different models of uncertainty, including several discussed above (e.g., Thomson, et al., Smithson), but expand to include concepts specific to linguistic data such as polysemy, homonymy, hedges and modals, as well as external factors including worldviews influenced by cultural perceptions and traditions.

Auger and Roy touch on one other interesting aspect for this thesis: verb tense. While statements about events or states which took place in the past may, to some extent, be verified as true or untrue, descriptions of most future events are inherently uncertain. There will be a fuller discussion of this later in Chapter 6.

Their view of uncertainty in linguistic data, shown in Figure 2.12, is divided into *linguistic ambiguities* and *referential ambiguities*. According to the authors' descriptions, linguistic ambiguity is tied to the meaning of the words (symbols) used in a given language, while referential ambiguity refers to cultural or other contextual elements which affect the content of the communication.

Here one should note the sub-branch "contextual elements" includes at least one leaf which is specific to spoken as opposed to written natural language communication, namely *body language*. N.B.: from the context of its placement in the graphic, one assumes *mood* indicates the speaker's feelings rather than grammatical mood, the former is assumed, since modal verbs and tenses are included under *linguistic ambiguities* and all other contextual elements appear non-grammatical.

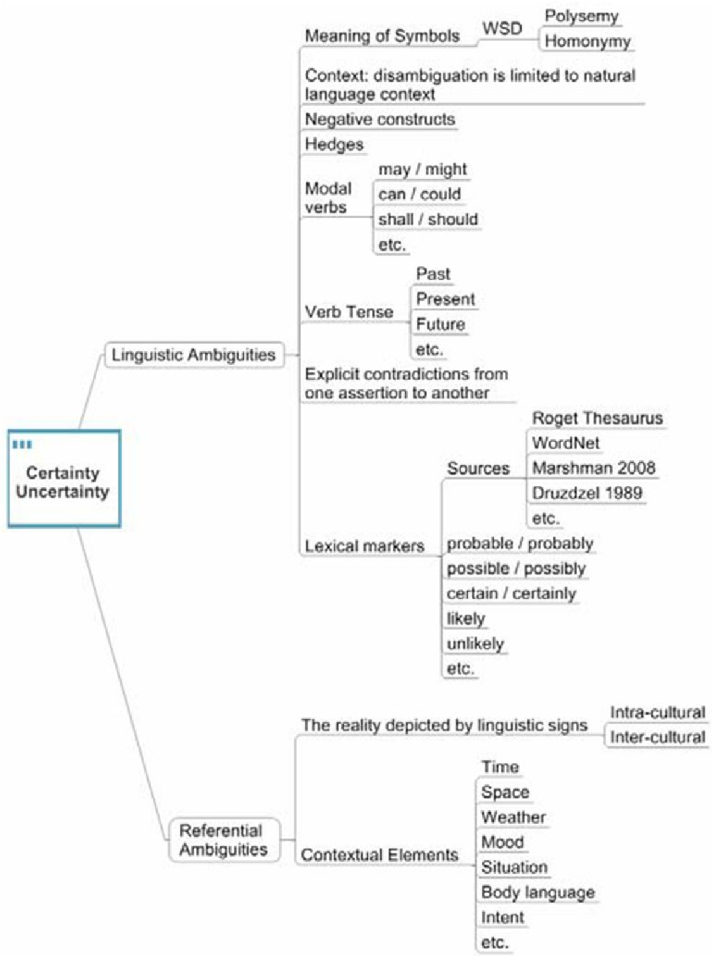


Figure 2.12: Auger and Roy divide certainty/uncertainty in linguistic data into two broad categories: linguistic ambiguities and referential ambiguities. [[Auger and Roy, 2008, p. 4]]

2.2.10 Uncertainty in Information Fusion

In this section we move closer to the application area in which lies the reason for this thesis: *uncertainty* in the area of information fusion. Kruger et al. [2008] examined the different types and levels of uncertainty in the information fusion process:

> There are two main levels of uncertainty in this process. The first level concerns each individual report or piece of information. This level is comprised of two parts:

- Source uncertainty: relative reliability of the information source, as adjudged either by the system (device sources) or evaluated by the reporter (human sources)
- Content uncertainty: the estimated veracity of the content of a report, assigned by the system (e.g., device sources) or as evaluated by the reporter (human sources)

> It should, of course, be noted that source and content credibility are generally not completely discrete. Particularly in the case of a human source, the reliability of the source has a direct impact on the credibility of the content: we tend to assign the information delivered by a reliable source a higher degree of credibility than the information we receive from someone whom we perceive to be unreliable. The second level of uncertainty concerns the interrelationship of various individual reports. This level is likewise comprised of two parts:

- Correlation uncertainty: this uncertainty results from the process of identifying and clustering potentially related reports, based upon the variances encountered in comparing features.
- Evidential uncertainty: result of matching reports to schemata which describe specific threats or situations. This measure of uncertainty is in many ways cumulative: its value is calculated based upon the values of source, content and correlation uncertainties from each of its mapped elements.

[Kruger et al., 2008, p. 686]

It is clear that this thesis is focused on *content uncertainty* as described above. As briefly discussed in Chapter 1, the main motivation behind the work in this thesis lie in finding a way to automatically generate weights which provide insight into how trustworthy propostions expressed in natural language may be based upon information in the sentence itself, rather than relying on the evaluation by an analyst to assign a weight.

2.3 Uncertainty Focus Within This Thesis

As demonstrated by the preceding overview, there are significant differences in how uncertainty is perceived depending upon the context and application of the model the authors were dealing with. Each of the models presented offers interesting viewpoints; however (unsurprisingly), none of the models above exactly suits the purposes of this thesis.

The focus of this research is upon written text – more specifically, written text at the sentence or sub-sentence level – and upon the (automated) exploitation of lexical and grammatical clues embedded in those sentences to assign an evidential value to the propositions which they contain for the purpose of building knowledge bases or for use in applications such as information fusion. Therefore, clues not contained in the text, such as body language, tone of voice, situational context, precision, and many of the other concepts which appear in models above are of no interest. We confine our definitions of uncertainty solely to the written words before us, and that only at the level of the sentence or below.

There are two basic categories of detectable uncertainty which appear at the sentence level:

- Inexactness, which concerns uncertainty within the propositional of the sentence, including imprecision, vagueness and ambiguity, and
- Evidentiality, which is the uncertainty about the propositional content of the sentence, including modal verbs, modal adverbs (including "words of estimative probability"), markers of hearsay, belief, inference, assumption, etc., and, in English to a certain degree, passive voice. (Note: selection of *evidentiality* as used in this sense is discussed in depth in Chapter 4.)

These are shown in Figure 2.13.

The sentence elements which express, directly or indirectly, uncertainty about the propositional content of the sentence can generally be easily identified and exploited by computer to provide an initial assessment of the certainty of the information which we receive in the form of written text. For example, we can use verb tense to differentiate between events or states which have occurred in the past as opposed to events or states which may or may not take place at some point in the future (which is uncertain until it

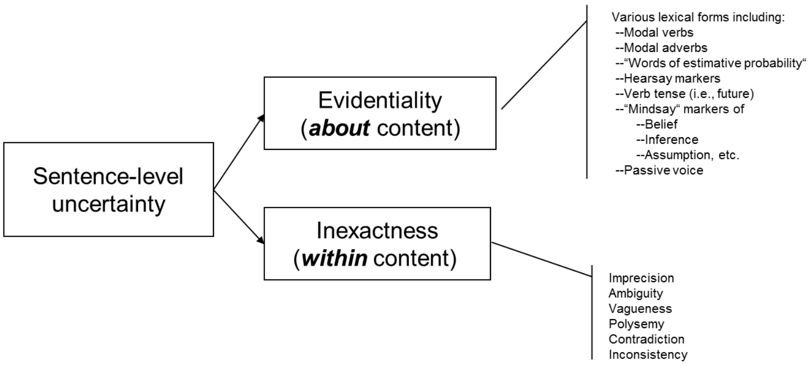


Figure 2.13: Uncertainty at the sentence level.

has happened). Other words or phrases also give us clear signals about the uncertainty of the: modal verbs (*might*, *could*, etc.), lexical markers such as modal adjectives (*possibly*, *likely*, etc.), or indicators of hearsay or opinion (*Sources said*, *I believe*, etc.).

The following chapter will explain both types of sentence level uncertainty in more detail, and demonstrate how this model was derived.

Chapter 3

Uncertainty in Natural Language

"...natural language sentences will very often be neither true, nor false, nor nonsensical but rather true to a certain extent and false to a certain extent, true in certain respects and false in other respects"

George Lakoff [1973, p. 458]

"Without hedging, the world is purely propositional, a rigid (rather dull) place where things either are the case or not. With a hedging system language is rendered more flexible and the world more suitable.

John Skelton [1968, p. 38]

3.1 Overview

At the conclusion of the preceding chapter, we presented a schematic in which we defined uncertainty in natural language at the sentence level with two main sub-types, *evidential* (uncertainty about the content of the proposition) and *inexactness* (uncertainty within the content of the proposition).

The distinction between these types is significant to our ends in this work. In particular, we are interested in the lexical and grammatical markers which signal the presence of potentially uncertain content. Therefore, we will explain this differentiation in more detail.

3.2 Uncertainty within the content

Natural language information may be uncertain in a variety of ways. Some of these are identifiable at the sentence level (e.g., *imprecision*, *vagueness*, *ambiguity*), while others are more generally found in larger contexts, i.e., over multiple sentences, or documents. Examples of the latter are *inconsistency* (i.e., conflicting, overlapping, gapping or confusing information), *inaccuracy* (incorrect information) and *contradiction*. As our work focuses on sentence-level analysis, we will confine the discussion below to those manifestations of uncertainty which can be found at the sentence level.

Uncertainty on the sentence level may manifest itself by vagueness or imprecision. That is, a sentence may contain elements which lack clarity or specificity, by using words which are open to interpretation. Consider, for example, sentence (8) which contains imprecise, non-specific information in several forms.

(8) *There were some animals in the road.*

Some expresses an imprecise number. The reader might possibly make some judgments on the range of numbers represented by *some*, although this may be bounded by other equally imprecise values such as *a couple*, *a few*, or *several*. Strictly speaking *a couple* refers to the quantity *two*. However, in colloquial usage, *a couple* may be used to indicate more than two (but not less). Cambridge Dictionaries Online [retrieved on Feb. 16, 2016] defines the venacular understanding in informal situations as "two or a few things that are similar or the same, or two or a few people who are in some way connected," while (retrieved on Feb. 16, 2016a offers the idiomatic interpretation: "a couple of, more than two, but not many, of; a small number of; a few." Merriam-Webster retrieved on Feb. 16, 2016a defines the informal usage of *a couple* as " an indefinite small number ... few" and Oxford Dictionaries(retrieved on Feb. 16, 2016a] defines *a couple* as an "indefinite small number." Therefore, had the writer used *a couple* rather than *some*, we would have been able to assume the number of animals was a quite small number two, three or, perhaps, four animals (but most definitely more than one). On the other hand, *many* would have been used if there were a noticeably larger number, such as twenty or fifty. *Several* is again somewhat slippery: more than *a couple* – "more than two but not many" Oxford Dictionaries [retrieved on Feb. 16, 2016b], "more than one or two, but not a great many" Cambridge Dictionaries Online [retrieved on Feb. 16, 2016] with Merriam-Webster (retrieved on Feb. 16, 2016b confusingly defining as "more than one" and "more than two but fewer than many" (the differentiation lies in the context). What thing is clear: *several* is less than *many*, which the various sources agree to be "a large number."

But it is not just the number of animals which is uncertain: the animals themselves are a second problem. The reader has no idea from the what sort of animals these may be: cats? dogs? cows? elephants? Or possibly there was a mixture of different types (a sheepdog and five sheep, for example). One type of animal would most likely not be applicable here: a human. In such a case, *people* or *persons* would have been used in place of *animals*. However, statement (8) does not necessarily leave out the presence of a human: for example, when a herd of sheep are in the road, the shepherd is usually somewhere in the vicinity as well, but the animals, not the shepherd, would likely be considered as some sort of anomaly worth mentioning. Similarly, a high level of precise detail may also be inaccurate: even if we were told that there were six brown Jersey cows in the road, it may well be that the observer neglected to let us know that there were two black herding dogs as well, or failed to detected that one brown cow was in fact a Hereford and not a Jersey.

Another vague element in the sentence is the road the animals were in. From the conversational context, that is, from the preceding communications, the listener may in fact know precisely which road is under discussion. Without that context, the reader would be unable to make such a determination and therefore the information remains vague. It is also entirely possible that this information is unimportant or unnecessary for the listener:

(9) *Sorry I was late arriving. There were some animals in the road.*

By proving a context in the form of the sentence preceding the original assertion as seen in (9), it is clear that the speaker is using (8) to explain the reason for his late arrival. Unless the listener knows absolutely that the speaker is arriving from somewhere else in the same road, the information conveyed is that somewhere along the route to the destination, which may have involved multiple roads, there was a road in which animals were present (and presumably posing a traffic hindrance).

Statements may be ambiguous, i.e., it may be open to more than one interpretation or have more than one possible meaning:

(10) *I saw her duck.*

(11) *Students hate annoying professors.*

In (10) it is unclear whether the female person referred to has dipped her head to avoid, say, a flying object or a low hanging branch, or whether she keeps a waterfowl as a pet. In (11) either the students dislike professors who irritate them, or whether students try to avoid making their professors angry, perhaps in the hope of receiving a better grade in the course.

The presence of vagueness, imprecision or ambiguity does not make the veracity of information which is conveyed in the assertion uncertain. In all of the above-mentioned examples, we have no indications from the sentence structure alone that the assertion is in anyway untrue or doubtful. We might have background knowledge which leads us to doubt the veracity of a statement because it contradicts other information that we have. Possibly we may have reservations about the reliability of the source of that information: for example, if we revisit (9) with the background knowledge that the speaker is chronically tardy and always has a prepared excuse to explain his tardiness, we might well doubt that the animals in the road existed, assuming that this assertion provides the speaker with a convenient, but unverifiable, explanation. However, without any such context, we have no reason not to believe the assertions above, regardless of whether we may not be quite certain what they exactly mean.

Thus, referring back to our model of uncertainty at the sentence level from the end of the previous chapter, the examples above illustrate the uncertainty within the content of the sentence, but there is nothing telling us we should doubt the truth of the statement.

But human communication often does not consist solely of the transfer of historical, factual information from one person to another. We humans make assumptions, express doubt or disagreement; we pass on information which others have told us, express our beliefs and wishes, make conjectures and assumptions, discuss events which may happen in the future (or might have happened in the past). We convey conclusions about the state of things based upon inferences made from other information. We may state our ignorance, or contradict statements of others.

Even when a statement contains a detailed, unambiguous and precisely described assertion, there may be clues that the speaker provides us with that indicate there may be some reason to not accept this assertion as the absolute truth. For example, the sentence may be formulated using a modal adverb or a modal verb form which introduces an element of doubt, or there may be other sentence elements which provide clues which indicate that the assertion is not the product of direct observation by the speaker, but derived by other means, leaving the veracity of the assertion open to some doubt. The following section describes in more detail, how these formulations allow us to make judgments about the truth of the proposition contained in the sentence.

3.3 Uncertainty about the content

Revisiting and modifying sentence (1) from the previous section, we provide some examples below which illustrate some of the ways in which the speaker provides us with information that we can use to decide how much to believe the proposition in the sentence.

(12) *There may have been some animals in the road.*

In (12), the modal verb expression *may have been* expresses either an element of the writer's doubt in the veracity of this statement or is an indication of speculation. While we have no information as to how to precisely interpret the writer's intent when using the modal verb, we do know that the writer is uncertain about the truth of the content of the assertion.

The modal adverb in (13) conveys the possibility of there having been animals but leaves some doubt:

(13) *It is possible there were some animals in the road.*

Whereas in (12) and (13), there is speculation or doubt about animals in the road that may have originated with the writer, in (14) we have a slightly different situation:

(14) *It was reported there were some animals in the road.*

Here, the writer indicates that the information comes from another source (i.e., hearsay) with no further clues as to where the information originated from. We have the choice of either accepting this as simply an accurate passing on of true information, or we can see this statement as uncertain, due to either aspects of the informational content may change during re-telling, or the possibility that there had been uncertainty clues in the original which were ignored or simply not passed on. Additionally, as we have no further information as to the source of the information (who reported?), we cannot make a judgment as to its reliability based upon background information we may have on that source. However, one can make the case that the writer would not have included a reference to a third party if she did not feel it important to convey that she herself had not witnessed the event and therefore, is distancing herself from the assertion, rendering it (at least slightly) uncertain.

It is a different case in (15) below:

(15) *Mary told me there were some animals in the road.*

Similarly to (14), in (15) the writer indicates that the information comes from another source. This time, however, we have more information as to where the information came from (Mary). Again, the writer distances herself by telling us the information came from another source, but, should we have some background knowledge about Mary's reliability as a source of information, we may adjust our belief on the veracity of the information based upon what we know about Mary.

In (16) we again have a formulation which indicates both an unidentified third party as the source of information (via *supposedly*), which additionally conveys a certain lack of confidence on the part of the writer regarding the truth of the proposition:

(16) *Supposedly there were some animals in the road.*

For example, the speaker may be passing on to the tardy colleague's excuse for his late arrival and wishes to convey that he does not quite believe the tale. (When delivered verbally, an accompanying eye roll or skeptical tone of voice would give the listener a more accurate idea of how much or little the writer believes this excuse; without these clues, we can only just the written word as expressing some level of doubt.)

Similar to the examples above, example (17) may convey hearsay but also something else:

(17) *I believe there were some animals in the road.*

The use of *I believe* in (17) may be interpreted in two ways: as an expression of belief on the part of the speaker that the tale is true, or possibly as an indication of hearsay ("Why was he late?" "I believe there were some animals in the road.").

From the context of surrounding sentences, we should be able to determine which usage is intended. For example, going back to the example of the oft-tardy colleague:

(18) *I know he often tells fibs, but not in this case. I believe there were some animals in the road.*

The first sentence in (18) sets the context which lends credibility to the second.

In (19) it becomes clear the correct interpretation is hearsay, in which the speaker passes on what she understands was the cause of an automobile accident.

(19) *There was a report on the news of a terrible automobile accident. I believe there were some animals in the road.*

However, regardless whether we can identify the statement as belief or hearsay (or a combination of the two) based on a broader context than just the single sentence itself, the assertion remains uncertain.

The above are, of course, just a few examples of expressions of uncertainty at the sentence level, and relatively simplistic ones at that. A quite exhaustive discussion of some of the more subtle manifestations of uncertainty may be found in Lindley [2006].

In the following sections we will examine in more detail markers and indicators of uncertainty in English sentences with the specific focus on uncertainty about the proposition rather than within it. Along the way we will see, as in the previous chapter about *uncertainty* as a concept, that there are as many different viewpoints and definitions as there are researchers.

3.4 Hedges, boosters, downtoners and other creatures

Very often when one is asked to consider markers of uncertainty in natural language, the first that comes to mind are modal adverbs: *possibly*, *probably*, *likely*, etc. The next categories are often modal verbs: *might*, *could*, *may*, etc., followed by nouns such as *likelihood*, *possibility*, *probability*, and so on. Lexical verbs like *suggest*, *assume*, *seem*, *guess*, etc., likewise convey uncertainty, as do adjectives such as *possible*, *probable*, *doubtful*, etc.

For many researchers, all of the above manifestations of uncertainty are included in a larger the category known as *hedges*. Oxford Dictionaries (retrieved on Feb. 16, 2016c defines this usage as a "word or phrase used to avoid overprecise commitment", Dictionary.com (retrieved on Feb. 16, 2016b) lacks the noun form, but as a verb describes *to hedge* as "to avoid a rigid commitment by qualifying or modifying a position so as to permit withdrawal." Merriam-Webster (retrieved on Feb. 16, 2016c) defines *hedge* with a slightly ominous twist: as "a calculatedly noncommittal or evasive statement."

The use of the moniker *hedge* is attributed to Lakoff [1973], who used this term to mean any lexical or grammatical form which indicates "fuzziness" in natural language. Using the mathematical theories of Lotfi Zadeh (the intellectual grandfather of imprecise mathematical theory, discussed previously in Chapter 2) as a basis, he defines a broad spectrum of lexical and grammatical elements in natural language which indicate any softening of the formulation of propositions, such that they express vagueness or imprecision:

> For me, some of the most interesting questions are raised by the study of words whose meaning implicitly involves fuzziness – words whose job is to make things fuzzier or less fuzzy. I will refer to such words as 'hedges'. [Lakoff, 1973, p. 471]

Figure 3.1 below shows a list of some elements that he considers hedges; it should be noted that since his focus in this initial paper was on category membership (prototype theory), *related phenomena* in this context appear to be such elements as *a real* or *a regular*, the usage of which emphasizes the strength of membership in a classification (e.g., *a real hero*); these forms would most likely not considered hedges now.

SOME HEDGES AND RELATED PHENOMENA	
sort of	in a real sense
kind of	in an important sense
loosely speaking	in a way
more or less	mutatis mutandis
on the ______ side (tall, fat, etc.)	in a manner of speaking
roughly	details aside
pretty (much)	so to say
relatively	a veritable
somewhat	a true
rather	a real
mostly	a regular
technically	virtually
strictly speaking	all but technically
essentially	practically
in essence	all but a
basically	anything but a
principally	a self-styled
particularly	nominally
par excellence	he calls himself a ...
largely	in name only
for the most part	actually
very	really
especially	(he as much as ...
exceptionally	-like
quintessential(ly)	-ish
literally	can be looked upon as
often	can be viewed as
more of a ______ than anything else	pseudo-
almost	crypto-
typically/typical	(he's) another (Caruso/Lincoln/ Babe Ruth/...)
as it were	______ is the ______ of ______
in a sense	(e.g., America is the Roman Empire of the modern world, Chomsky is the DeGaulle of Linguistics, etc.)
in one sense	

Figure 3.1: Lakoff's list of hedges and related phenomena [1973, p. 472]

Since Lakoff's first article, the definition of hedging has shifted to focus on expressions of uncertainty or commitment on the part of the speakers. However, as with many areas of research, there is not necessarily complete agreement on what *hedge* means. Below we describe a few of the various viewpoints. Holmes [1982] does not specifically use the term *hedge* but focuses on *epistemic modality* and, in particular, looks at the major grammatical classes which are generally now considered to be hedges such as modal verbs, lexical verbs, adverbial constructions, etc. She also defines the two terms *boosters* and *downtoners*:

> In general, lexical devices used to express strong conviction can be described as Boosters, whereas those used to signal the speaker's

> lack of confidence or to assert something tentatively can be characterized as Downtoners. Boosters strengthen or increase the illocutionary force of utterances while Downtoners weaken or reduce their force.[Holmes, 1982, p. 18]

Prince et al. [1982] look at hedges more in the original sense of Lakoff and classify them in two main categories: *relational hedges* that have to do with the speaker's relation to the propositional content, and *propositional hedges* that introduce uncertainty into the propositional content itself. They further break the two types of hedges into four different types: (1) *rounders*, which show approximate ranges for quantitative information (e.g., about 10, roughly a dozen); (2) *adaptors*, which suggest the similarity of non-identical cases (e.g., sort of, a kind of); (3) *plausibility shields*, when the speaker is not fully committed to the assertion or the assertion is not based on deductive logic but plausible reason (e.g., *seems like*, *appears to be*); and (4) *attribution shields*, when the speaker attributes the assertion to another person or object (e.g., *according to*).

Chafe and Nichols [1986] defines hedges more narrowly as one of several categories of *evidentials*, which he defines quite broadly as "any linguistic expression or attitude toward knowledge" (see Chapter 4 for a fuller discussion of evidentials):

- reliability of knowledge: maybe, probably, surely, undoubtedly;
- knowledge as having been arrived at through some kind of reasoning, e.g., obvious, must, should
- knowledge as having been derived from a particular kind of evidence (sensory evidence or hearsay. e.g., see , hear, feel, it sounds like, it seems, supposed to be
- hedges: sort of, kind of, etc.

Biber [1988] sees hedges in a broader scope and picks up on Holmes' idea of downtoners but with a slightly different twist:

> Hedges are informal, less specific markers of probability or uncertainty. Downtoners give some indication of the degree of uncertainty; hedges simply mark a proposition as uncertain. [Biber, 1988, p. 240]

For Hyland [1998], with a particular focus on scientific writing:

> ..."hedging" refers to any linguistic means used to indicate either a) a lack of complete commitment to the truth value of an accompanying proposition, or b) a desire not to express that commitment categorically. [Hyland, 1998, p .1]

Hyland uses the term *booster*, in the original sense from Holmes, as a counterpoint to hedging but has replaced her *downtoner* with *hedge*. According to Hyland, boosters "allow writers to project a credible image of authority, decisiveness and conviction in their views. This definition is shared by Vázquez and Giner Vazquez and Giner [2009], who follow Hyland's lead, arguing that boosters are used to support persuasion.

In his 2000 paper, Hyland provides some examples of boosters and hedges that he used in the course of his study as shown in Figure 3.2:

Table 1: Boosters and Hedges used in the study.

Target Items in Questionnaire	
Boosters	show that/always; demonstrate/substantially; clearly show/will; fact that; obviously/will
Hedges	suggest /may; seem; believe/could; appear to; might; hypothesise; assume /likely; speculate; possible; might;
Target Items in text and questions	
Boosters	clearly show (4); clear (3); definite (2); certain; fact that (2); show/always (1)
Hedges	might (5); possible (3); may (3); suggest (2); seem (2); hypothesise; likely (2); speculate; believe/could; assume; probably; indicate (1)

Figure 3.2: Examples of boosters and hedges from Hyland's study [2000, p. 184]

Hyland also points out that hedges may be used for politeness or deference; in other words, they may have social meaning rather than being indicators of uncertainty:

> Hedges such as might, probably and seem signal a tentative assessment of referential information and convey collegial respect for the views of colleagues. Boosters like clearly, obviously and of course allow writers to express convictions and to mark their involvement and solidarity with an audience. [Hyland, 2000, p. 179]

Furthermore, Hyland [1994] includes several other phenomena in his definition of hedging, including passive voice, conditionals (if-clauses), question forms, impersonal phrasing and time reference. Particularly in scientific writing, the use of passive voice and impersonal phrasing are widely, almost universally, used, conveying an undertone of "but I might be wrong or have overlooked something." With regard to impersonal phrasing, Hyland writes,

> ... 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.[Hyland, 1994, p. 240]

Clemen [1998] defines hedging thus:

> Die Hecke (das Hedging) ist ein interaktionales, der Diskursanalyse zuzuordnendes Sprachmittel in der gesprochenen und geschriebenen Kommunikation. Sie hat eine pragmatisch-kommunikative Funktion und erlaubt dem Sprecher/Schreiber,
>
> - seine Aussagen zu subjektivieren
> - seine Verantwortung für den Wahrheitsgehalt der Proposition zu relativieren
> - den Grad seiner Gewißheit oder seines Zweifels über die Geltung einer Feststellung einzuschränken
> - absolute Aussagen zu vermeiden
> - Verantwortung für Äußerungsinhalte zu transferieren
> - persönliche Einstellungen zu bekunden und Sachverhalte zu bewerten,
>
> womit er sich Rückzugsmöglichkeiten verschafft, Unsicherheiten verbergen kann, das Risiko des Irrtums minimiert, einen potentiellen Einwand des Rezipienten antizipiert und das interpersonale Kommunikationsverhältnis optimiert. [Clemen, 1998, pp. 14-15]

She expands further:

> Der sprachlichen Fixierung der Hecke dienen im wesentlichen lexikalische Mittel aus dem Modalwort-, Modalverben und Modalpartikelbereich, epistemische Operatoren, abmildernde (more fuzzy)- und als einstellungsspezifische Indikatoren bekräftigende (less fuzzy)-Markierungen, reservierte, vage und reduzierte Verantwortungsübernahme ausdrückende Lexeme und Wendungen, Geltungseinschränkung bewirkende hypothetische Formulierungen, hypothetische Notwendigkeit ausdrückende Konstruktionen mit performativen Verben und einstellungsbekundende Adverbien und Adjektive. [Clemen, 1998, pp. 15]

Vassileva [2001] looks at hedging in terms of commitment and detachment of the writer and includes "classes of boosters termed as 'solidarity' (the case when the author claims shared knowledge with the audience) and 'belief' (when the author states unequivocally that he/she is absolutely convinced of what he/she is saying)."

Pappas [1989] describes hedges as belonging to *qualifiers*, that is, they are indicators of the level of approximation and speaker commitment to the main assertion. Under this description, *probably*, *appear*, *partially*, or *a tendency to* are classified as hedges, while Pappas uses the term *intensifier* where other authors use *booster*, namely, as elements which reflect speaker confidence.

In her work exploring linguistics patterns which express uncertainty in conceptual relations, Marshman [2008] separates modal verbs (*can*, *could*, *may*, *might*, *should*, etc.), from hedges (*more or less*, *roughly*, *somewhat*, *mostly*, *essentially*, *very*, *especially*, *exceptionally*, *often*, *almost*, *practically*, *actually*, *really*, etc.). Interestingly, she also designates negative constructions as being a source of uncertainty.

Crompton [1997] defines a hedge as "an item of language which a speaker uses to explicitly qualify his/her lack of commitment to the truth of a proposition he/she utters." In an early attempt to find common ground as to what hedges are, he examined the viewpoints of various authors regarding hedges, and then proposed the following "characterisations of hedged propositions":

1. Sentences with copulas other than *be*.
2. Sentences with modals used epistemically.

3. Sentences with clauses relating to the probability of the subsequent proposition being true.
4. Sentences containing sentence adverbials which relate to the probability of the proposition being true.
5. Sentences containing reported propositions where the author(s) can be taken to be responsible for any tentativeness in the verbal group, or non-use of factive reporting verbs such as "show", "demonstrate", "prove". These fall into two sub-types:
   (a) where the authors explicitly designate themselves as responsible for the proposition being reported;
   (b) where authors use an impersonal subject but the agent is intended to be understood as themselves.
6. Sentences containing a reported proposition that a hypothesized entity X exists and the author(s) can be taken to be responsible for making the hypothesis. [Crompton, 1997, p. 284]

Like Hyland, Crompton is focused on academic and scientific writing, commenting,

> It seems that there is a danger of hedge being used as a catch-all term for an assortment of features noticed in academic writing. Clearly, the use of impersonal construction, passivization, lexis expressing personal involvement, other politeness strategies, and factivity in reporting/evaluating the claims of other researchers are important issues in academic writing; these all seem worthy of further research to enhance the teaching of the subject. However, the restriction of hedge to designate language avoiding commitment, a use which corresponds closely, as we have seen, with the ordinary use of the word, seems desirable and feasible, both theoretically and pedagogically. [Crompton, 1997, p. 286]

Prokofieva and Hirschberg [retrieved Sep 15, 2015] define hedges as "speculative cues" which "can be a single word or a combination of multiple words that signal uncertainty, a lack of precision or non-specificity, or an attempt to soften or downplay the force of the speaker's utterance." They also point out that not all speculative language is necessarily a hedge, using hypotheticals as examples of a non-hedging speculation. However, they also make the point that conditional or speculative formulations are not always to be considered as hedging:

> Hypotheticals such as 'If it rains, I won't go to the game' contain instances of speculative language (if/then) but are not considered instances of hedging behavior. [Prokofieva and Hirschberg, retrieved Sep 15, 2015, p. 1]

In order to identify hedge cues, they propose asking three questions as a "litmus test":

- Is the speaker being deliberately uninformative (or under-informative)?
- Is the speaker uncertain?
- Is the speaker trying to downplay the force of their utterance? [Prokofieva and Hirschberg, retrieved Sep 15, 2015, p. 1]

Since the focus of their paper is on the annotation of hedges in text, they provide very detailed explanations on the identification of various types of hedges and their scope: *multi-word*, *negated*, *disfluent* and hedges in questions with a focus on application.

While not using the specific linguistic terms such as "hedge", Liddy et al. [2004] have developed a conceptual, four-level model of certainty developed with a focus on reported information (newspapers, journals, etc.), which captures a number of the concepts in the preceding discussion. In particular, their three contextual dimensions (later refined upon by Rubin [April 2007]) cover topics of interest for us. Dimension 2 ("Perspective") differentiates between the writer's point of view and third party information; third party perspective differentiates between that of participants (e.g., direct witnesses) and experts (e.g., subject matter experts).

Dimension 3 ("Focus") essentially separates fact from less reliable content such as hearsay, opinion, etc. As with Hyland, they include time (tense) as a fourth dimension also contributes: while future events may be of interest to us, until they have become past events they are uncertain: "I may state I will go to the conference next month, but that is before I end up sick in the hospital. I might regularly attend a meeting every Wednesday, but next

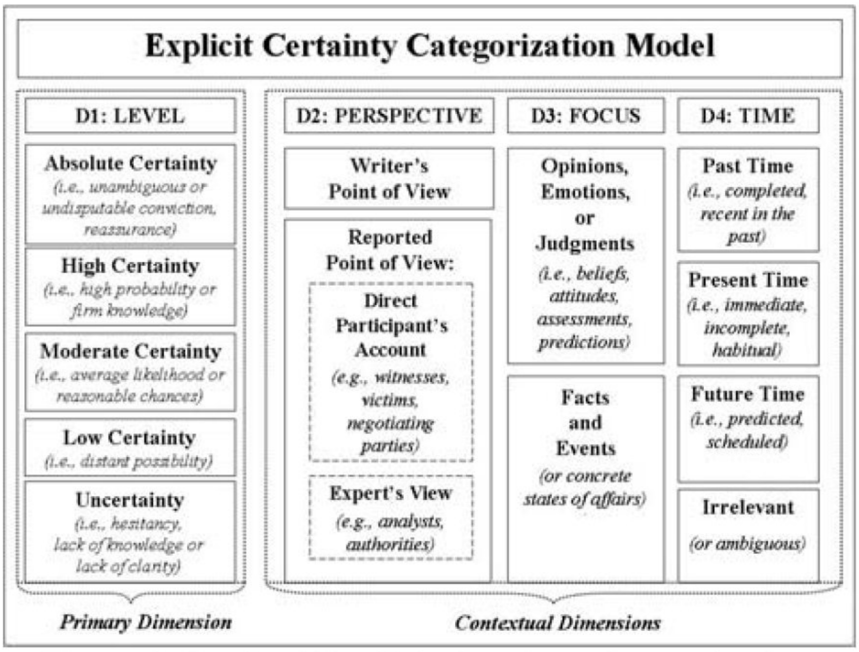


Figure 3.3: Explicit certainty categorization model based upon reported information model from [Rubin, April 2007, p. 142], expanding upon initial work by Liddy et al. [2004]

week I have to travel, so I will not be there." Depending upon the application using the information which is being extracted, future events, or projection based upon habitual behavior may be of great interest (for example, if the application is predictive) – or of no interest whatsoever. In either case, tense may play a role.

3.5 Uncertainty for this thesis

It is clear from the above that we are once again confronted with the same words being used in different ways by various authors, sometimes varying only by degree and sometimes varying fairly dramatically. Some authors focus on both content-based hedges (e.g., some, few, many), where others focus only on those constructs which reflect on the strength or weakness of the commitment of the writer to the proposition. Some researchers write of boosters and downtoners, others of boosters and hedges. Some researchers include elements such as verb tense and voice, hearsay and mindsay, and politeness strategies.

For our purposes within this thesis, we will define the following:

- *Hedges* are sentence elements which:
    1. reflect the reliability of knowledge, including adverbs (e.g., *maybe*, *probably*, *surely*, *undoubtedly*), verbs (*should*, *might*, *appear*), adjectives (*possible*, *likely*), nouns (*possibility*, *likelihood*)
    2. flag the knowledge as having been arrived at through some kind of reasoning, (e.g., *assume*, *infer*, *conclude*, *must*), or
    3. flag the knowledge as having been derived from a particular kind of sensory evidence or hearsay, (e.g., *see*, *hear*, *feel*, *it sounds like*, *it seems*, *supposed to be*).
- *Boosters* are lexical elements which intensify ("boost") the hedge. For example, adding the booster *very* to *likely* increases the strength of the original.
- *Downtoners* are lexical elements which weaken the hedge. For example, adding the downtoner *somewhat* to *likely* will weaken the original.

It should be noted here that direct observation in the first person is not considered a hedge: unless we doubt the source – only tangientially an aspect of the work in this thesis – we take direct first person observation at its face value. However, reports of direct observation by other individuals would be considered hearsay.

Exactly how these concepts will work together will be discussed and demonstrated at length in Chapter 7.

Chapter 4

Evidentiality, Epistemic Modality, or Epistemic Stance?

"When I use a word,' Humpty Dumpty said in rather a scornful tone, 'it means just what I choose it to mean — neither more nor less.'
'The question is,' said Alice, 'whether you can make words mean so many different things.'
'The question is,' said Humpty Dumpty, 'which is to be master — that's all."

Lewis Carroll, *Through the Looking Glass* [1872, p. 72]

What's in a name? That which we call a rose by any other name would smell as sweet.

William Shakespeare , *Romeo and Juliet*

4.1 Overview

At the end of the previous chapter, we indicated that the focus of the work in this thesis would be on the various lexical elements contained in a sentence which provide us clues as to the veracity of the propositional content in that sentence. One of the important tasks when proposing a new methodology or concept is the careful selection of the nomenclature which will be used to describe its components. The first step in such a selection is the examination of existing terms to determine their applicability to the new concept. In some happy instances, there is an existing moniker which fits perfectly. However, it often happens that there is no immediately apparent candidate. There may even be disagreement among researchers as to how to classify related concepts; for some, the various elements may be classified in discrete, separate categories, while others may be lumped together in different configurations. Even when the clustering may be virtually identical, it may not necessarily appear under the same designator. Thus, as happens in such cases, one must discuss the various alternatives and determine the appropriate terminology for the topic at hand.

In this chapter we will examine a trinity of related concepts: *epistemic modality*, *evidentiality* and *epistemic stance*. Depending on the focus and background of the researcher, these terms may be applied differently: for some researchers, these terms may be considered to be three separate (but often closely related) concepts, for others, these are three names for more or less the same thing.

Particularly for English, which, as we discuss below, contains none of the specialized grammatical forms expressing evidentiality existing in a number of other natural languages such as Tuyaca from Columbia (Barnes [1984]) or Quechua from Peru (Weber [1986]), relying instead on a lexically-based "evidential strategy" to accomplish this purpose, it turns out that these three names can be viewed as three facets of the same phenomenon. This phenomenon is central to this work, and hence it is worth investigating these topics in some depth. This discussion will provide the basis for the selection of terminology which will be used during the remainder of this thesis.

4.2 Epistemic modality

Fintel [2012] defines modality as "a category of linguistic meaning having to do with the expression of possibility and necessity." Modality may be expressed in a variety of ways in English: modal adverbs (*possibly*, *probably*), modal verb auxiliaries (*should*, *must*), semi-modal verb forms (*ought to*, *need to*), adjectives (*possible*, *likely*), nouns (*possibility*, *likelihood*), and constructions such as conditionals (*if... then*).

Epistemic modality is the subcategory which expresses possibility or necessity based upon the knowledge of the speaker. [Halliday, 1970, p. 349] describes it as "the speaker's assessment of probability and predictability. It is external to the content, being a part of the attitude taken up by the speaker: his attitude, in this case, towards his own speech role as 'declarer'." Bybee and Fleischmann [1995] and Palmer [1986], among others, view epistemic modality as indicating the speaker's commitment to the proposition. This has often been interpreted as meaning that epistemic modality (as well as other types of modality) represents the speaker's belief as to the veracity of the proposition.

As stated above, epistemic modality expresses possibility or necessity based upon the knowledge of the speaker, it is not surprising that hints about where the speaker has derived the knowledge may be implicit in many forms of epistemic modality. For example, in the statement "It's 7 o'clock, John might be home now", the use of "might" could be seen as indicative of assumption or inference on the part of the speaker. This interpretation of "might" can be viewed as a slide into the direction of "evidentiality," which, narrowly defined, is an indication of the source of the information contained in the proposition as defined by Aikhenvald [2004]. Taking a wider view of evidentiality, Westmoreland [1998] argues that "may" and "might" should not be considered as modals expressing necessity and possibility at all but should be considered as evidentials. [Rooryck, 2001b, p. 166] agrees that evidentiality may "piggyback" on other constructs – which he dubs "invisible" evidentiality – and provides some examples of this using modals verbs in German and Dutch as markers of hearsay:

(20) *Es soll bisher vier Tote gegeben haben.*

(21) *Jan zou in het geheim naar Brazilië geëmigreerd zijn.*

This blurring of the boundaries and inconsistent interpretation between epistemic modality and evidentiality is not unusual. For example, SIL's LinguaLinks defines epistemic modality as "a modality that connotes how much certainty or evidence a speaker has for the proposition expressed by his or her utterance" [retrieved Jun 10, 2015]. Further, SIL classifies "evidentiality" and "judgmental modality" as subsets of epistemic modality – "judgmental modality" being defined by them as "an epistemic modality that connotes the speaker's strength of inference, or degree of confidence in the reality of the proposition expressed by his or her utterance." However, the definition SIL offers for evidentiality is "an epistemic modality that connotes the speaker's assessment of the evidence for his or her statement. An evidential is a form, such as a verbal affix, that is a grammatical expression of evidentiality." By restricting this definition to "a form, such as a verbal affix", the number of languages in which "evidentiality" appears is reduced, and, more importantly, English is not in this group. Therefore, one could argue that, under this definition, evidentiality simply does not exist in English, and, as a result, there is only epistemic modality.

According to Nuckolls and Michael [2014]:

> ...we can probably attribute the conflation of evidentiality and epistemic modality that was characteristic of early work on evidentiality to the fact that speakers of languages that lacked grammaticalized evidentials found it difficult to understand evidentials as anything other than a proxy for epistemic modality, which was a familiar category to them.[Nuckolls and Michael, 2014, pp. 13-14]

DeHaan [1999] sees the two concepts as completely distinct:

> While the literature on the subject makes it appear at first glance obvious that evidentiality and epistemic modality are closely related, there is just as much evidence, if not more, to cast serious doubt on this analysis. It is not the case that evidentiality is a subcategory of epistemic modality. Rather, we are dealing with two distinct categories: one, evidentiality, deals with the evidence the speaker has for his or her statement, while the other, epistemic modality, evaluates the speaker's statement and assigns it a commitment value. [DeHaan, 1999, p. 25]

But, as we will see in the following section, it turns out, there is not only disagreement about whether evidentiality is or is not a subset of epistemic modality (or indeed vice versa), there is also disagreement as to exactly what evidentiality is or is not. The following section gives an overview of discussions on this topic.

4.3 Evidentiality

Despite their morphological similarity, the linguistic concept of "evidentiality" is not necessarily related to the idea of "evidence" in the usual, common sense of the word. Oxford English Dictionary defines evidence as "the available body of facts or information indicating whether a belief or proposition is true or valid" retrieved Apr. 4, 2015. Websters defines it as "[t]hat which makes evident or manifest; that which furnishes, or tends to furnish, proof; any mode of proof; the ground of belief or judgement; as, the evidence of our senses; evidence of the truth or falsehood of a statement" retrieved Apr. 4, 2015 and WordNet's definition includes "Your basis for belief or disbelief; knowledge on which to base belief" retrieved Apr. 4, 2015. Rather, it is generally agreed that evidentiality has to do with the source of the information conveyed in the proposition and, depending on who you read, also on the speaker's certainty about that information. Just as there is discussion about whether evidentiality is a type of epistemic modality or not as described above, there is a discussion as to what does, or does not, belong to the concept "evidentiality."

The study of evidentiality in languages first appeared in the middle of the 20th century. Boas [1947] is credited with the first mention of this concept in his discussion of the Kwakiutl language, in which he identifies "a small group of suffixes express[ing] the source and certainty of knowledge". He used "evidential" to describe one of the suffixes in the group. It was, however, Jakobson [1957] who gave the first definition of evidential: "a tentative label for a verbal category which takes into account three elements – a narrated event, a speech event and a narrated speech event... namely the alleged source of the information about the narrated event." Jakobson recognized four sources of information: retelling another person's narration (hearsay or quotative evidence), a dream (relative evidence), a guess (presumptive evidence) and one's own experience (memory). There is no conveyance of "proof" of the veracity or lack thereof in the assertion being made. Rather, evidentiality identifies where the knowledge contained in the statement has come from. Depending upon the author and/or language in question, there are different types of evidentiality. Willett [1988] defined several types of evidentiality. Aikhenvald expanded upon his definitions while ignoring others to produce a slightly different list.

For the purposes of this thesis, we have collapsed the various types of evidentiality identified by different researchers into three basic categories:

- Direct: first person sensory information, which is sometimes separated into visual and non-visual (auditory, tactile, taste). Some authors such as Aikhenvald separate these into two separate categories.
- Hearsay: reported evidence from other sources and which includes second person and third person information. Under *second person*, the writer repeats information from a direct source. Under *second person* Aikhenvald and others define a more explicit form of hearsay is sometimes referred to as "quotative" or "mediated", in which the information source is explicitly named. In contrast, in *third person hearsay*, the writer repeats knowledge from indirect sources which may include general knowledge, "the grapevine" or even gossip. Willett includes *folklore* as an evidential, which might be interpreted as "general knowledge" shared by a group of individuals with a given culture. Anderson [anderson:86] includes "general reputation", and "myths and history."
- "Mindsay": a term taken from Bednarek [bednarek:06] which was coined in contrast to "hearsay", as the knowledge is produced by reasoning such as inference, in which a conclusion is based on reasoning using tangible or visual evidence or intangibles such as logic, general knowledge, etc., as well as personal or emotional components such as belief, conjecture or assumption. It should be noted that hearsay may contain mindsay, in that the writer might report (hearsay) the opinion (mindsay) of another person.

Over time, the concept of evidentiality slowly became of increasing interest to scholars, and, not surprisingly, began to be defined in different ways.

There are essentially two schools of evidentiality. The first school, represented most strongly by Aikhenvald [2004], sees evidentiality as simply the expression of the nature of the source of the knowledge. In particular Aikhenvald and others in this camp focus predominantly on the grammatical forms used to convey this source information. While only a minority of natural languages contain such forms, many indigenous languages in the Americas (such as the previously mentioned examples from South America, Tuyaca and Quechua) have quite sophisticated grammatical systems for expressing information about the source of information, including hierarchies of trustworthiness of the sources. As we have seen from the SIL example in the previous section, some researchers belonging to this group believe that the languages without grammatical forms have no evidentiality, others (including Aikhenvald) believe that such languages are capable of expressing evidentiality via evidential strategies which use lexical elements and other formulations to identify the source of the knowledge. English belongs to those languages which have no grammatical forms for the expression of evidentiality, so it relies on the usage of various lexical elements to represent this information.

The second school of evidentiality considers that evidentiality conveys not only information about the source of the content of the proposition but also expresses the speaker's judgment about the reliability of that content. According to Chafe and Nichols [1986], "[e]vidential markers are defined as grammatical categories which indicate how and to what extent speakers stand for the truth of the statements they make."

Rooryck [2001a] expands upon this:

> Evidentials indicate both source and reliability of the information. They put in perspective or evaluate the truth value of a sentence with respect to the source of the information contained in the sentence, and with respect to the degree to which this truth can be verified or justified. This justification can be expressed by markers referring to immediate evidence on the basis of visual observation, to inference on the basis of (non)observable facts, to deduction or inference, etc." [Rooryck, 2001a, p. 125]

Similarly, in the oft-cited definition by Anderson [Anderson, 1986, p. 274] it is said that evidentials give "the kind of justification for a factual claim which is available to the person making that claim" and this indication of evidence has to be the primary meaning of the evidential structure. Nevertheless, not all linguists agree with this one and only meaning given to evidentiality. On the one hand there is a consensus that "[t]he basic characteristic of linguistic evidentiality is the explicit encoding of a source of information or knowledge (i.e. evidence) which the speaker claims to have made use of for producing the primary proposition of the utterance" [Diewald and Smirnova, 2010, p. 1]. According to Willett [1988], this view corresponds to evidentiality in the narrow sense, because an explicit relationship between evidentiality and modality is denied.

For Biber and Finegan [1989], Ifantidou [1986] and others, the identification of source and the writer's assessment of the reliability of the knowledge expressed in the statement are considered intertwined.

DeHaan [1999] disagrees with this linkage:

> ...epistemic modality and evidentiality both deal with evidence, but they differ in what they do with that evidence. Epistemic modality evaluates evidence and on the basis of this evaluation assigns a confidence measure to the speaker's utterance. This utterance can be high, diminished, or low. An epistemic modal will be used to reflect this degree of confidence. An evidential asserts that there is evidence for the speaker's utterance but refuses to interpret the evidence in any way."[DeHaan, 1999, p. 4]

He adds further:

> ...evaluation is obviously done on the basis of evidence (which may or may not be expressed overtly, or which may or may not be expressed by means of evidentials), but there is nothing inherent in evidentials that would compel us to assign an a priori epistemic commitment to the evidence."[DeHaan, 1999, p. 25]

While this statement may be applicable in the case of languages in which evidentials are grammatical elements, the inclusion of information concerning the source in languages such as English with no (mandatory) grammatical evidential forms can be regarded to be a significant comment on the speaker's judgment regarding the proposition. In other words, since there is no grammatical requirement to include this information, one can assume that the speaker had a reason to do so. Frajzyngier [1985] states "the different manners of acquiring knowledge [that evidentials denote] correspond to different degrees of certainty about the truth of the proposition." In a language in which such information is added voluntarily, this may be taken as a clue to the commitment of the speaker. Only through interpretation of the form in which this information has been presented will we be able to assess what the speaker's intent in adding this information has been: to distance himself from the proposition, to provide more authority (in order to convince the listener), to wish to remain neutral.

The discussion in this section clearly demonstrates that there is disagreement among researchers as to what does or does not constitute evidentiality.

Once again, we are presented with no clear guidance as to the perfect classification for our purposes, although many of the arguments, including those of Rooryck [2001a], in particular but also Biber and Finegan [1989], Ifantidou [1986], and Anderson [1986] support the inclusion of both the (original) source of and the writer's assessment of the reliability of the information under *evidentialitiy*.

In the following section, we investigate *epistemic stance* as a third and final alternative concept.

4.4 Epistemic Stance

Stance is a broader concept than either evidentiality or epistemic modality – indeed some argue that stance subsumes them both, as well as other constructs which we have discussed in previous chapters. Precht [2003], for example, claims stance subsumes hedges, evidentiality, vague language, attitude, affect and modality. Nordquist [retrieved Jun 13, 2015] defines stance more loosely as "[l]inguistic and non-linguistic forms and strategies that show a speaker's commitment to the status of the information that he or she is providing."

Biber and Finegan [1989] define stance as "the lexical and grammatical expression of attitudes, feelings, judgments or commitments concerning the propositional content of a message." They examined a variety of characteristics that fall under their definition of stance, identifying six interpretive types, which stance characteristics are most distinctive to each of these types, and in which varieties of text these distinctive characteristics are most likely to be found. Their analysis of stance characteristics is shown in Figure 4.1.

"The linguistic expression of attitude", the authors continue, "has been studied under two main topics: evidentiality and affect." [Biber and Finegan, 1989, p. 92] They define *evidentiality* following Chafe's inclusion model, i.e., as reflecting the writer's "expressed attitudes towards knowledge: towards its reliability, the mode of knowing and the adequacy of its linguistic expression."[Chafe and Nichols, 1986, p. 271] *Affect*, in contrast, reflects personal attitudes on behalf of the writer, including emotions, moods, expectations and (non-epistemic) judgments such as surprise or disappointment.

Interpretive Label	Distinctive Stance Characteristics	Major Text Varieties
1) 'Emphatic Expression of Affect'	+ + + affect markers + + emphatics, hedges, certainty verbs, doubt verbs, possibility modals	+ + + personal letters + conversation
2) 'Faceless Stance'	+ + + absence of all stance features	+ + + written exposition + + written fiction + + informational spoken texts
3) 'Interactional Evidentiality'	+ + + hedges, emphatics + + certainty and doubt adverbs, certainty and doubt verbs, possibility modals	+ + + conversation
4) 'Expository Expression of Doubt'	+ + + doubt adjectives + doubt adverbs	+ + informational written exposition
5) 'Predictive Persuasion'	+ + + certainty adjectives + + predictive modals	+ + letters of recommendation
6) 'Oral Controversial Persuasion'	+ + predictive modals + certainty adverbs, possibility modals	+ + informational spoken texts + formal letters

Figure 4.1: Stance styles according to Biber and Finegan 1989, p. 116

Some researchers extend stance to cover other phenomena in human communications. Fairclough [1992] considers that stance markers may also define relationships between writer and reader, including defining the power hierarchy or indicating solidarity. For example, an expression of uncertainty may not necessarily reflect uncertainty about the content of the proposition on the part of the writer, but rather his timidity as the weaker partner in the communication.

Marin-Arrese [2009] focuses on a subset of stance, which she calls epistemic stance:

> Epistemic stance refers to the knowledge of the speaker/writer regarding the realization of the event and/or to his/her assessment of the validity of the proposition designating the event. Linguistic resources for the expression of the various forms of stance include modal, evidential and attitudinal expressions. [Marin-Arrese, 2009, p. 23]

She further subdivides epistemic stance into three parts: epistemic modality, personal evidentiality and mediated evidentiality, and, interestingly for us, assigns reliability classifications to several examples, as shown in Figure 4.2 below.

Epistemic modality	Personal evidentiality		Mediated evidentiality	
Implicit Subjective	Explicit Subjective	Explicit/Implicit Intersubjective	Subjective Attribution: Endorsement/ Distancing	Objective Attribution: Evidential standing of Source
High Reliability: Certainty *must P* *Certainly P*	**High Validity** *I saw P* *I know P* *I say to you P*	**High Validity** *We have witnessed P* *We all know P* *Clearly P*	**High Validity: Overt/Covert Endorsement** *X has compellingly argued P* *X has revealed P*	**High Validity: Authority or High social standing of Source** *Scientists have discovered that P* *Everybody knows P*
Medium Reliability: Probability *will P* *Probably P*	**Medium Validity** *I think P*	**Medium Validity** *It seems P* *You would think P* *Apparently P*	**Medium Validity: Covert distancing** X has claimed P	**Medium Validity** *X has said P* *Reportedly P* *It is believed P*
Low Reliability: Possibility *may P* *Perhaps P*	**Low Validity** *I feel P* *I suppose P*	**Low Validity** *You get the feeling P* *Supposedly P*	**Low Validity: Overt Distancing** X has mistakenly held the view P	**Low Validity** *Some people believe P*

Figure 4.2: Epistemic stance according to Marin-Arrese 2011, p. 793

The concept of mediated evidentiality, that is, indirect evidentiality via secondary sources, is particularly of interest. The strategy of using (and often naming) other external sources to provide strength to an argument, thereby "legitimizing" both the proposition and the resulting argumentation, is widely used. Indeed, this thesis itself is filled with such references; without these external citations, the value of the results would be called severely into question. White [2006] refers to the practicing of using experts, prestigious social status or widely accepted information as sources as giving the proposition "evidential standing." Interestingly enough, this shifts the burden of the truth of the proposition elsewhere, although the speaker may invoke any of the many other means at her disposal to represent her commitment to the credibility of the proposition.

Once again, we are reminded of the interconnectedness of all three concepts on which we have focused, while at the same time it is clear that there is no consensus among researchers as to their meanings.

However, we need a single term for that which we are investigating in this thesis. Therefore, the following section discusses possible alternatives, and proposes a solution.

4.5 And the winner is....

The focus of this thesis is on the analysis of lexical items which indicate uncertainty about the truth of the propositional content of an utterance. For the purposes of this thesis, in order to provide cohesion and avoid unnecessarily redundant references, we need to either define a new term entirely or to select one of the three which have been discussed above.

Before we begin this final discussion, we should note two important points:

- While one may fundamentally view the work of this thesis as trying to determine the "truth" of a proposition by examining linguistic clues contained within a sentence, it should be viewed more subtly. There are a number of communications which are neither true nor false, for example, when the event being described has not yet happened, where we wish to make a judgment as to whether we believe it might be true at some (later) point. Since the results of this work will be used to support decision-making under uncertainty, we are trying to assess the strength or weakness of the proposition, rather than its truth, based upon information the writer has embedded within the sentence.

- The work being presented in this thesis, while oriented toward application, in particular, in the area of information fusion, and therefore, the consumers of the results of this research may well come from non-linguistic backgrounds; thus a compromise between that which may satisfy a linguist and that which makes sense to, say, a computer scientist or a modeler is desirable. Therefore, while various researchers in linguistics may disagree with this choice of nomenclature, we ask them to keep this point in mind.

These points having been made, we may proceed with our discussion of options for the selection of an appropriate descriptor for our purposes.

One of the possibilities is to examine descriptors other than those listed above as a candidate. Many words are already in use within the area of application for which this work is designated. For example, as we will discuss in Chapter 5, the descriptor *belief function* has a specific, well understood and widely used meaning in the area of applied mathematics called Dempster-Shafer which is very widely used in the domain of information fusion. Since the results of the work done in this thesis will very likely flow into mathematical models based upon this mathematical system, it would be confusing for practitioners. It likewise makes little or no sense to assign new meaning to such a term (or even its abbreviated form *belief*.

Similarly, while words such as *credibility* and *reliability* spring to mind as likely candidates, they too are problematic within one major area of application (intelligence) for which this work is intended. In the intelligence community, *reliability* refers solely to the source of information, and is represented by six pre-assigned values which all users understand. Similarly, *credibility* is used with the intelligence community to designate the "truthfulness" of the information content that a source has delivered; while this appears to be the ideal designator for our work, *credibility* likewise has six pre-assigned values (one of which is essentially “cannot make a judgment”) which practitioners in the field use. Furthermore, as we have discussed in Chapter 3, uncertainty about the propositional content of a sentence relies both upon the source of that content, as well as other lexical markers of uncertainty with which the source surrounds the proposition, using one of these terms to represent both concepts (source and content uncertainty) would be confusing for practitioners. Therefore, these two terms are also suboptimal for our purposes.

However, returning to the three concepts presented earlier in this chapter, there is one term which would fit both of the criteria mentioned above: being understandable for non-linguist consumers, but likewise acceptable to linguists. This candidate is ***evidentiality***.

From the non-linguist point of view, the strongest argument for evidentiality is its morphological similarity to *evidence*, a word which is commonly understood among English speakers. Re-examining the definitions of evidence discussed at the beginning of Section 4.2, and adding a new one, we find that certain elements of these definitions apply perfectly to the first of our points above:

- *the ground of belief or judgement* [Webster's Online Dictionary retrieved Apr. 4, 2015]

- *basis for belief or disbelief* [WordNet retrieved Apr. 4, 2015]
- *ground for belief* [American Heritage Dictionary retrieved on Aug. 23, 2014]

While linguists disagree as to what does or does not fall under the category *evidentiality*, there are a number of researchers (e.g., Chafe and Nichols, Rooryck, Biber and Finegan, Ifantidou) who support the idea that, to once again quote Rooryck,

> [e]vidential markers are defined as grammatical categories which indicate how and to what extent speakers stand for the truth of the statements they make. Evidentials illustrate the type of justification for a claim that is available to the person making that claim. [Rooryck, 2001a, p.125]

Therefore, in light of the conditions which we outlined at the beginning of this section, and the ensuing discussion, as well as the foreseen application of the results of this thesis, we have opted to select *evidentiality* for use within this thesis.

Chapter 5

Quantifying Uncertainty

A reasonable probability is the only certainty.

E.W. Howe [1926, p. 23]

Million to one chances crop up nine times out of ten.

Terry Pratchett *Mort* [1987]

5.1 Introduction

Identifying and understanding the types and sources of uncertainty present in textual information is only one step toward using this information for decision making. Most models using this information are based upon mathematical algorithms to calculate the reliability of the results. Thus, it is not enough to simply identify that a statement is uncertain and in which way it is uncertain, we also need to assign some sort of (quantitative) value to each piece of information which reflects our perception of its accuracy.

Over the past centuries, mathematicians, philosophers and logicians have attempted to create mathematical models to represent and – even more importantly – predict uncertainty and risk in the real world. One of the first significant models for uncertainty has been Bayesian Probability, the foundations of which go back to Thomas Bayes in the seventeenth century. The principles of Bayesian probabilities originally was created to support betting hypotheses, predictions of the outcomes of rolling dice or of drawing a particular hand of cards.

Until the twentieth century, the Bayesian view of the world was expanded upon, but never essentially departed from. However, with the explosion of research in discrete and applied mathematics, and based upon new needs driven by developments such as computer science and artificial intelligence, a number of new theories and models for the representation of uncertainty arose.

In 1965 things changed when Lofti Zadeh [1965] presented an extension to classical set theory called fuzzy sets. In classical set theory, the membership of an element in a given set was defined in a "crisp" binary, "either-or," "true/false" fashion: the element was a member of that set or it was not. However, in real life, boundaries are often not as clear-cut: elements can "sort of" belong to a set. Depending upon their characteristics, some elements may be held to be more representative of a given set than other elements. Less representative elements may have membership functions in more than one set, because they lie in boundary regions and "kind of" belong to more than one set.

There are numerous approaches to mathematically formalizing uncertainty. In Figure 5.1, Klir and Smith [2001] divide various theories into additive (classical numerical probability), and nonadditive (everything else). Kohlas and Monney [1994] further subdivide the nonadditive theories into non-standard probability theories and non-probability models. Non-standard probability theories include, but are not limited to, the Dempster-Shafer theory of evidence (Shafer [1976]), the Hints models (Kohlas and Monney [1995]), the upper and lower probabilities models (developed by numerous researchers including Good [1950] and C.A.B. Smith [1961], etc.). From the concepts underlying fuzzy set theory sprang a number of non-probabilistic mathematical models. In [1978], Zadeh extended his fuzzy set theory into possibility theory, which was itself later further extended into belief functions by Dubois and Prade [2001] and the transferable belief model by Smets and Kennes [1994].

UNCERTAINTY THEORIES			FORMALIZED LANGUAGES				
			CLASSICAL SETS	NONCLASSICAL SETS			
				FUZZY SETS	ROUGH SETS	FUZZY ROUGH SETS	ROUGH FUZZY SETS
MONOTONE MEASURES	ADDITIVE	CLASSICAL NUMERICAL PROBABILITY	1	2			
	NONADDITIVE	POSSIBILITY/ NECESSITY	3	4			
		BELIEF/ PLAUSIBILITY	5	6			
		CAPACITIES OF VARIOUS ORDERS	7				
		COHERENT LOWER AND UPPER PROBABILITIES	8				
		CLOSED CONVEX SETS OF PROBABILITY DISTRIBUTIONS	9				
		. . .					

Figure 5.1: Classification of uncertainty theories [Klir and Smith, 2001, p. 10]

Finally, although not a mathematical theory per se, we will discuss the representation of uncertainty by odds, an alternate representation of probabilities used by statisticians and probabilists, as well as gamblers (although there are some differences in implementations). This representation along with the linguistic formulations it generates are often used both by laypersons and experts to express uncertainty. As a result, it is worth a short excursion into this more informal representation.

5.2 Probability theories

Probability theory is one of the two classical theories of uncertainty, arising in the mid-17th century through the works of mathematicians such as Pascal, Fermat, Huygens and through actuarial work of deWitt, Hudde and

Graunt Smets [July 2, 1999]. For some two hundred years, probability theory remained the sole theory for representation and calculation of uncertainties, extended and further by numerous mathematicians including Euler, Laplace and Bayes. During the 20th century, new needs led to the development of competing theories, some, such as possibility theory, were based upon probability theory, others such as fuzzy sets broke new ground.

Probability theory is well-documented as a methodology for formalizing, and calculating uncertainties, with a plethora of texts describing the basics. However, though many lay people tend to use "probability" in the sense of a single concept, there are in fact four sub-theories which describe different situations. These are discussed in depth in Smithson [1993].

In each case the basic premises of probability theory remain intact:

- There exists a finite set of mutually exclusive outcomes.
- Each outcome is assigned a specific likelihood and the sum of those likelihoods within the set of outcomes is 1.0 .

Probability theory and its variations continue to play a major role in the evaluation of risk. One major reason for this is that there are comprehensive axiomatic foundations for these theories. A second reason is tractability: a numerical result can be quickly and easily determined using these methods.

5.2.1 Classical probability theory

The original definition of probability theory was based upon the idea of a set of equally possible events – the flip of an (unbiased) coin, the tossing of a (fair) die, or drawing a playing card from a (complete and unmarked) deck – in which the likelihood of a single given outcome was derived by the division of 1 (the sum of all outcomes) by the number of elements, i.e., 1/6 for each possible outcome of the throw of a die whose faces each contain a unique symbol. The "universe" of the set is known, the set of outcomes is known.

5.2.2 Relative frequency theory

In this variation of probability theory, the elements within the set of outcomes are not assumed to be equally likely. Each element of the set is assigned a likelihood based upon its relative frequency as established through numerous independently repeated trials. That is, the likelihood is not based upon absolute knowledge as in the classical theory, but upon heuristics developed through observation over time. In other words, occurrences of the desired event in question relative to all events will be counted over time and used as a basis to determine the probability of the desired event occurring in the future will be determined.

5.2.3 Subjective (Bayesian or personal) probability

In contrast to the two preceding theories, this variant of probability theory is not based upon either knowledge or observation of the universe, but based upon one's subjective belief that a specific outcome (event) will occur. Therefore, it is personal to the individual assigning the probability to the outcome, and hence probabilities arising from this are considered to be "degrees of belief" or an individual's personal opinion as to the probability of each event.

Subjective probability is widely used for the expression of opinion or interpretation. Very often the assignment of a probability to a possible future event is based upon a personal, sometimes emotional, not always rational belief as to the likelihood of that event occurring. For example, a sports fan can make assign a subjective probability to the likelihood of his favorite sports team winning the championship even before the season has begun, based not only on factors such as previous performance or changes in team membership but also on irrational factors such as "it's about time for them to win again." Thus the weighting of the belief about this outcome may vary widely from fan to fan, depending on the factors each individual uses to generate the probability.

5.2.4 Logical (a priori) probability

Whereas classical probability assigns probabilities of single events based upon complete knowledge of the universe under consideration (e.g., the faces of a die), the logical schools assign probabilities based upon the evidence at hand. Keynes [1962] defines *a priori probability* as a logical relation between a proposition and a corpus of evidence. Using logic rules we can derive a new hypothesis using information which we already know and our assumptions about this information. The probabilities which result are considered "logical" because they are entailed by (inductive) logic. In this sense, logical probability may therefore also be considered "objective," in contrast to the pure subjectivity of Bayesian (personal) probability. It should be noted, though, that because assumptions are involved, logical probabilities are necessarily "subjective" to a certain extent.

Regardless which variation used, one thing is common to all flavors of classical probability: a single precise value assigned to a single event. This single value also allows for additive manipulation of results to describe more complex behavior.

However, it has long been accepted that it is difficult, if not impossible or impractical, to force the assignment of a single value to events which are not so discrete. It is one thing to assign a single value to each potential outcome of a coin toss, it is another thing to assign a single value to events which do not have strict demarcations such "sick" or "partly cloudy". With the first, we have complete knowledge of all possible (discrete) outcomes – heads or tails – and can make predictions based upon these. In the determination of, say, the health of farm animals, the determination of the states "sick" and "healthy" are not so clear. Assigning a single (numerical) value is much more difficult, if not just downright impossible.

To summarize these succinctly, according to according to Hajek [2012]:

> Broadly speaking, there are arguably three main concepts of probability:
>
> - A quasi-logical concept, which is meant to measure objective evidential support relations. For example, "in light of the relevant seismological and geological data, it is probable that California will experience a major earthquake this decade".
> - The concept of an agent's degree of confidence, a graded belief. For example, "I am not sure that it will rain in Canberra this week, but it probably will."
> - An objective concept that applies to various systems in the world, independently of what anyone thinks. For example, "a particular radium atom will probably decay within 10,000 years.Hajek [2012]

In each of these statements, the "probable" can be associated with a distinct percentage: there is a 75% chance that California will experience a major earthquake, or a 30% chance that it will rain in Geneva. Or I may be 90% certain that my favorite team will go to the playoff finals this year.

However, in real life, it is seldom so clear cut. Even though we speak of a 75% likelihood, what we really mean is that it is much more likely that there will be a major earthquake in California than not. Just as fuzzy set theory was created to provide an alternative to the binary set allocation of classical set theory, it became necessary to build other mathematical theories and models to deal the messiness of real life. In the 20th century, mathematicians began to develop theories of imprecise probabilities that allow us to model the world using imprecise, uncertain and incomplete data.

5.3 Classification and Set Theories

Traditionally, we think of sets as groups or collections of elements which somehow belong together. This "belonging" together is based upon some sort of defined relationship which is defined by specific characteristics which the elements share. Based upon these characteristics things being examined are determined to belong or not to belong to a given set. One can consider such a set to be classification based upon specific characteristics shared by the elements of the set. In classical set theory, set membership is clearly defined – an element is either in a given set or it is not – and the interrelationships between various sets may be easily determined.

While this approach may work for many types of elements, there are many more instances where such an approach will not work. The lines between sets may be blurred, or the elements themselves are not clearly delineated. For example, the values red, orange and yellow are not clearly delineated: there may be questions as to the boundaries between the values (i.e., at what point does red become orange?). These boundaries between the concepts red and orange are vague; there is a certain arbitrariness to assigning any sort of crisp delineation between the two concepts, which is, to a great extent, dependent upon personal opinion. Labov's classic discussion 1973 of categories, prototypes and classification discuss this topic in great detail

There are two main approaches to dealing with vague membership in sets. The first, and most well-known, is fuzzy set theory. This was proposed by Zadeh in 1965 when he introduced a variation of set theory to deal with vagueness. The basis of his work relies on the concept that an element may belong only partially to a set. For example, a height measurement of 2 meters for a human being would be universally considered "tall"; however, does an individual of height 180 cm also fall into this category? Or one of height 175 cm? Is the latter more likely to be considered a member of the set "medium height"? The element under consideration will be assigned a value ranging between 0 and 1 which determines the degree to which that element belongs to a given set. The membership of an element at the union or intersection of more than one set is determined based upon its individual memberships in each of the sets under consideration.

Figure 5.2 illustrates how various fuzzy concepts such as few, many, and some can be mapped. While the arguments can be made that none and all should be viewed as crisp values (0% and 100% respectively), these graphics illustrate that none may in fact be almost none, and all may be nearly all.

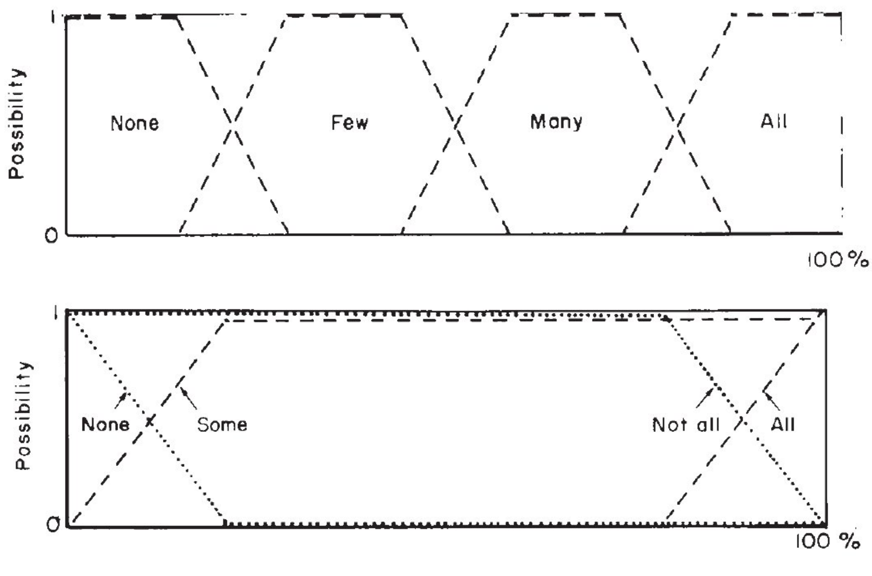


Figure 5.2: Representations of fuzzy concepts such as *few*, *many*, and *some* from [Zimmer, 1984, p. 126]

It should be noted that, because the value given for the degree of membership is between 0 and 1, it is sometimes erroneously viewed as the "probability" that the element belongs to this set, rather than a measure of the imprecision of the classification of the element to a specific set. In other words, the membership value 0.75 assigned to a given element does not mean that the element belongs 75% to this set and 25% to another, but rather that it is not as representative of the set as those elements with higher memberships (generally, it tends toward a boundary of the set).

It should also be noted that there are numerous ways in which membership functions may be assigned. Smithson [1993] has determined four approaches to this:

- Formalist: in which lower and upper bounds (0 and 1) for membership are agreed upon and defined solely in mathematical terms, with intermediate membership values defined by a smooth function;

- Probabilist: the degree a membership of an object is the (possibly subjective, possibly through polling) probability that it belongs to the set;

- Decision-theoretic: the degree of membership is defined by the utility, or payoff, of asserting that the object does indeed belong to the set;

- Fundamental measurement: numerical memberships assigned are quantitative which behave like fractional counts, based upon axiomatic conditions which can be shown empirically.

The second approach to imprecision and sets is rough set theory, which was proposed by Pawlak [1982]. Rough sets differ from fuzzy sets in that imprecision is expressed by a boundary region of the set, and not membership in the set. Rough set theory determines those elements which are inarguably in the set and those which are inarguably not in the set (the upper and lower approximations of the original set), and identifies a boundary region, which contains all of those elements which may not be definitively defined as in or not-in the set under consideration. If the boundary region is empty, then the set is considered to be crisp, i.e., there are no elements which cannot be clearly identified as either belonging to or not belonging to the set and therefore no ambiguity.

Intuitively one can see that for applications using natural language, a representation such as fuzzy or rough sets for imprecise concepts is a very useful tool. This is particularly so in that human beings often disagree on the classification of concepts; imprecise representation is excellent for reflecting this diversity.

In the following sections we will examine some of these theories briefly. Again, it must be noted that the purpose here is not to provide a rigorous mathematical representation of each theory presented, but rather to provide the reader with a "sampling" of some of the more important theories and their variants, which will function as a basis for later discussion within this paper.

5.4 Theories of Imprecise Probabilities

The field of imprecise probabilities is expanding rapidly. New theories emerge, while there are increasingly more variations, refinements and expansions of existing theories. In Figure 5.3 below, Klir and Smith [2001] offer a snapshot of imprecise probability theories ordered according to their levels of generality.

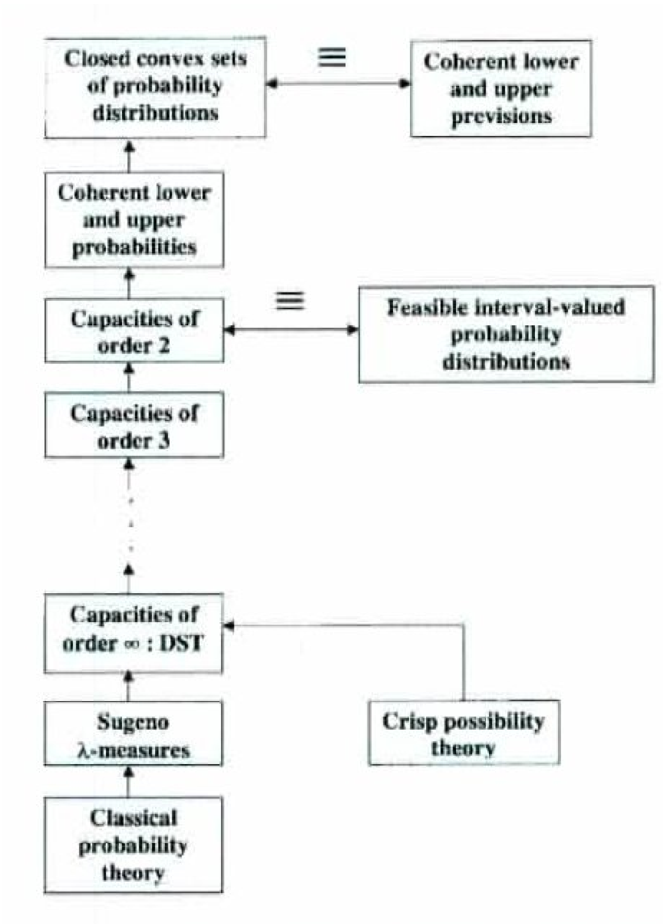


Figure 5.3: Classification of uncertainty theories from [Klir and Smith, 2001, p. 18]

This graphic gives some idea of the range of focus of these different theories. In the following subsections we look at several theories which are widely used.

5.4.1 Possibility Theory

Possibility theory is similar to probability theory, but is an extension of fuzzy set theory rather than classical set theory. The first reference to a "Theory of Possibility" was proposed by Zadeh [1978], who originally intended to use this for representing gradations in natural language formulations such as those used in fuzzy sets, although others such as the philosopher David Lewis and L.J. Cohen prepared much groundwork.

Dubois and Prade [2003] describe the concept of possibility with four characteristics:

- Feasibility: it is possible to do something (physical)
- Plausibility: it is possible that something occurs (epistemic)
- Consistency: compatible with what is known (logical)
- Permission: it is allowed to do something (deontic) [Dubois and Prade, 2003, slide 4]

In possibility theory the middle two characteristics are of interest: how plausible an event is, and how consistent it is with information that we already have. Therefore, where probability theory is based upon crisp propositions, that is, where the probability of a given event is represented by a single ("crisp") number, possibility theory is based upon two concepts: the possibility and the necessity of an event. The possibility of an event represents the extent to which the event is consistent with the information which we have, while the necessity expresses the extent to which the event is definitely implied by the knowledge, i.e., the extent to which our knowledge indicates "it must be so". The possibility measure provides an upper bound (everything that might be), while necessity provides a lower bound (everything that must be). Using these two measures allows us to represent partial belief and ignorance and therefore the ability to reason with imprecise probabilities Dubois et al. [2000].

5.4.2 Theory of Evidence (Dempster-Shafer)

The Dempster-Shafer theory of evidence (DST) can be viewed as a generalization of subjective probability (Shafer [1976]). Its strength lies in its assignment of probabilities to sets, thereby allowing it to represent ignorance. It is also able to cope with varying degrees of precision without difficulty. In fact, when the precision reaches the point that sets all become singletons (i.e., each contains only a single element), DST effectively collapses back into Bayesian probability.

The use of Dempster-Shafer is to test hypotheses for which there is at least some support from the available evidence. There are three central functions to DST. The first of these is the *basic probability assignment* (bpa), which refers not to the "classical" assignment of probability but rather defines the mapping of the power set (that is, all subsets of the set of hypotheses, including the empty set and the set itself) to the interval between 0 and 1 where the bpa of the empty set is 0, while the sum of the bpas of all subsets of the power set is 1. The second function is the *Belief function*, which represents the sum of the masses of all subsets that at least partially support the hypothesis being tested, thereby forming a lower bound. The third function is *Plausibility* which is 1 minus the sum of the masses of all subsets whose intersection with the hypothesis is empty (that is, they show no support at all for the hypothesis). This forms an upper bound.

Central to this theory is the set, called the frame of discernment, which contains all possible (distinct) values for the variable under consideration. While the elements in the frame of discernment may be numerical (generally an interval), they may often be other sorts of values. For example, the frame of discernment d for the variable h which represents the variable height may be defined as $d = \{$ tall, normal, short   in which the values are fuzzy representations rather than numbers.

Each of these elements will be assigned a numerical measure of belief (generally an interval) by a knowledgeable source (e.g., expert). Similarly, a value expressing a level of belief may be assigned to any subset of *d*. For example, we can define a subset *H*, where $H = \{$ tall, normal $\}$, that is given a belief equivalent to the statement "the value of *h* is tall or normal".

Under probability theory, in which the collective "beliefs" for a proposition must sum to 1.0 (100%), if a numerical value w is assigned to a proposition which is a subset of *d*, the remainder of one's belief (1 - w) is automatically assigned to the complement of that subset to assure the summing to 1.0. In Dempster-Shafer, one is allowed to leave this remainder uncommitted as an indication of lack of knowledge. This seems a trivial difference, but is important when combining evidence: "uncommitted" does not automatically act as evidence against the hypothesis. The strength of this approach is that it supports a certain intuitive approach to the handling of uncertainty, which has a certain pragmatic appeal. Furthermore, it supports the combination of evidence for a specific hypothesis; the Dempster Rule of combination is a powerful part of this theory.

Dempster-Shafer, it should be noted, is very widely used for the representation of uncertainty. There are a number of variations on the Dempster-Shafer theme, including Theory of Hints (Kohlas and Monney [1994]) as well as the Transferable Belief Model (Smets and Kennes [1994]), which is discussed briefly in the next section. The ability to combine uncertain evidence makes DST very interesting. However, researchers do not agree on how this should be done: there are a number of different variations including Yager [1986], Inagaki [1991] and Zhang [1994].

5.4.3 Transferable Belief Model

Smets and Kennes [1994] have proposed a variant to DST which utilizes belief functions, albeit in a different fashion than DST. The Transferable Belief Model is based upon the following:

- a two-level model: there is a credal level where beliefs are entertained and a pignistic level where beliefs are used to make decisions.
- at the credal level beliefs are quantified by belief functions.
- the credal level precedes the pignistic level in that, at any time, beliefs are entertained (and updated) at the credal level. The pignistic level appears only when a decision needs to be made.
- when a decision must be made, beliefs at the credal level induce a probability measure at the pignistic level, i.e. there is a pignistic transformation from belief functions to probability functions. [Smets and Kennes, 1994, p. 3]

Smets and Kennes claim that one significant difference from DST is that the Transferable Belief Model does not at any time reduce to a probabilistic model as does DST.

5.5 Odds

While not a formal mathematical theory similar those discussed above, odds are an alternative representation of relative probabilities. This may also be seen as the ratio of favorable and unfavorable outcomes. Odds are expressed as a ratio of two numbers in the form *x to y*, with alternate forms *x/y*, *x-y* or *x:y* in which *to* is substituted by a symbol. When this ratio is described as *odds in favor* the ratio represents the probability that the event will occur, while *odds against* describes the converse. Rather than stating, as in a previous example, that there is a 75% probability that California will experience a severe earthquake in the coming year, using odds we can represent this as 3 to 1 odds in favor of an earthquake. (*Odds for* and *odds on* are alternate expressions for the same thing). If there is a roughly 50% chance of occurrence, the expression *even odds* may be used to describe the probability, or, somewhat more often the description will be *a 50/50 chance* or *50/50 odds*.

Odds often feel more natural or are more convenient to use in certain instances. For example, describing the probability of a lottery win as 1:460,000 is more understandable for most people, than the corresponding probability expressed as a decimal with many zeros behind the decimal point. As a result, the representation of probability has been integrated into everyday language. In fact, in his original article Sherman Kent (1964) anecdotally relates the use of odds by his analyst colleagues (see Chapter 1).

5.6 Conclusion

While this is not a complete overview of mathematical representations of uncertainty, it does demonstrate that there are essentially two categories of representations – those which deal with crisp (discrete) values and those which deal with fuzzy values – and that within each of these categories there are various systems which are currently being used by researchers and practitioners. Ideally, we would find a solution to our problem which could be represented under both classes of systems. It turns out that our model will in fact accommodate both.

The next steps are to look at the work which has been done up to now in quantifying – that is, assigning numerical values to – the uncertainty which appears in the domain of natural language utterances, as well to identify and fill the gaps that still exist for which quantification has not yet been developed.

Chapter 6

Quantifying Evidentiality in English

Solum ut inter ista vel certum sit nihil esse certi.

Pliny the Elder, *The Complete Works* [2015, p. 130]

When one admits that nothing is certain one must, I think, also admit that some things are much more nearly certain than others.

Bertrand Russell, [1949, p. 115]

6.1 Introduction

In Chapter 3 we narrowed our scope down to hedges – all those elements within a sentence which convey some information as to how the main informational content was derived and which indicate the commitment of the speaker to that information content – our focus now turns to how to assign to that content numerical weights representing the evidential value of that content, which can then be used in algorithms.

[Schrage, February 20,2005, p.B01] notes

> [a] growing number of fields ranging from medical diagnostics to Internet spam filtering [. . .] increasingly rely upon Bayesian analysis – a probability theory that predicts the likelihood of future events based on knowledge of prior events – as a powerful tool to weigh new evidence.

Bayesian (subjective) probability, as we have seen in the previous chapter, however, is only one of many mathematical theories which are used to represent and combine uncertain information.

In this chapter we take a look at some of the work which has been done thus far in assigning numerical values to evidential expressions. As we will see below, this earlier work tends to focus almost exclusively on certain types of hedges, in particular expressions containing modal verbs (*should*, *could*, *might*), other verbs indicating conviction or another source (*believe*, *doubt*, *according to*, *assume*, *guess*), adverbs *(possibly*, *probably*, *likely*), adjectives (*it is possible*, *probable*) and some nouns (*possibility*, *likelihood*). In most of these studies, attempts have been made to ascertain numerical "values" for the various expressions, generally by asking participants in a study to locate the expression along a scale, from, say, 0 to 100 or to assign a percentage. Furthermore, hedges, as noted in Chapter 3, are often strengthened or weakened by the use of boosters and downtoners (*very likely*, *rather improbable*), requiring appropriate adjustmenst to their assigned values.

Other types of hedges, for example, those dealing with informational source such hearsay, conjecture, inference, etc., may often be ranked relative to each other in a hierarchical sense, but there appear to have been no attempts to assign numerical values to the evidentials in this category.

To make things even more complex, hedges do not always appear alone in a sentence: *I believe it is possible that Mary could be. . .* ; Clausen found "that uncertain sentences often contained multiple hedge cues, sometimes up to 4 or more."[2010, p. 124] Therefore, we often must try to assign a weight to the proposition which is based upon the interaction of multiple hedges.

One solution is to determine, in advance, all possible combinations of hedges, boosters and downtoners and assign them individual values. This would be a brittle solution, broken as soon as a combination does not appear in the table of values. Luckily, Crompton [1997, p. 284 points out that "compounding of hedges is quite common, but the elements of each compound are still distinguishable"; the reader can easily corroborate Crompton's assertion in the example from the preceding paragraph (*I believe it is possible that Mary could be...*). Recognizing this provides us with a basis to support a more robust solution, which is to determine the weights assigned to individual hedges and assign a composite weighting.

However, as we will see, ultimately the numerical values per se are irrelevant, in so far as there are no definitive "universally true" numerical values assigned to any of these hedges. Indeed, in the examples that follow one may clearly see that various researchers have used their own different (arbitrary) weighting scales, resulting different weight values and ranges for the same hedges.

What does, however, appear to be "universally true," as we shall see in the following sections, is the general ordering in which humans tend to organize the various hedges with their accompanying boosters and downtoners. This is of particular significance to us, as it allows us to assign "relative" weighting while at the same time freeing us from being tied to a specific mathematical weighting system. In other words, it allows us the freedom to assign, for example, single evidential value (crisp weights) or a range of values (fuzzy weights) to any given hedge or chain of hedges depending on the underlying application which is being used.

6.2 "Words of estimative probability"

There are many applications in which the processing of large volumes of unstructured text-based information is of great importance, among these are law enforcement, crisis management, business intelligence, state and national security, as well as the military in both conflict and non-conflict (e.g., peacekeeping) activities. In particular, as decisions made based upon such applications may be quite literally life-or-death, it is very important that the law enforcement, national security and military consumers of such gathered information be very focused on its reliability. Therefore, it should come as no surprise that much of the research in analyzing information quality comes from the national security and military side.

As briefly discussed in Chapter 1 (and repeated here for the convenience of the reader), in his 1964 article about what he refers to as "words of estimative probability" ("WEPs"), Sherman Kent [1964] of the United States Central Intelligence Agency relates the following anecdote about an intelligence report concerning the possibility of a Soviet invasion of Yugoslavia:

> 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'? What kind of odds did you have in mind?" 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. Understandably troubled by this want of communication, I began asking my own colleagues on the Board of National Estimates what odds they had had in mind when they agreed to that wording. It was another jolt to find 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. [Kent, 1964, p. 2]

What makes this anecdote particularly interesting is that the various individuals with whom Kent spoke were all individuals who were working in the same domain (intelligence), who most likely had similar educational backgrounds and, presumably, who also similar training for their analyst positions. In spite of all this, this anecdote shows us that hedges are open to interpretation.

Intrigued by this phenomenon, another CIA analyst Jr. [1999] performed an informal study and requested a number of colleagues to assign a single probability to a number of commonly used hedges. Figure 6.1 shows the hedges along with a mapping of the various probabilities assigned to each hedge.

The probabilities assigned to a number of the hedges were clustered very closely (*better than even*, *about even*, *highly unlikely*). A number varied quite dramatically: *highly likely* ranged across a span of more than 40 percentage points, as did *improbable*, *probably not* and *chances are slight*, while the range for *probable* started at 25% as the lower bound to just over 90% as the upper – a spread of more than 65 percentage points.

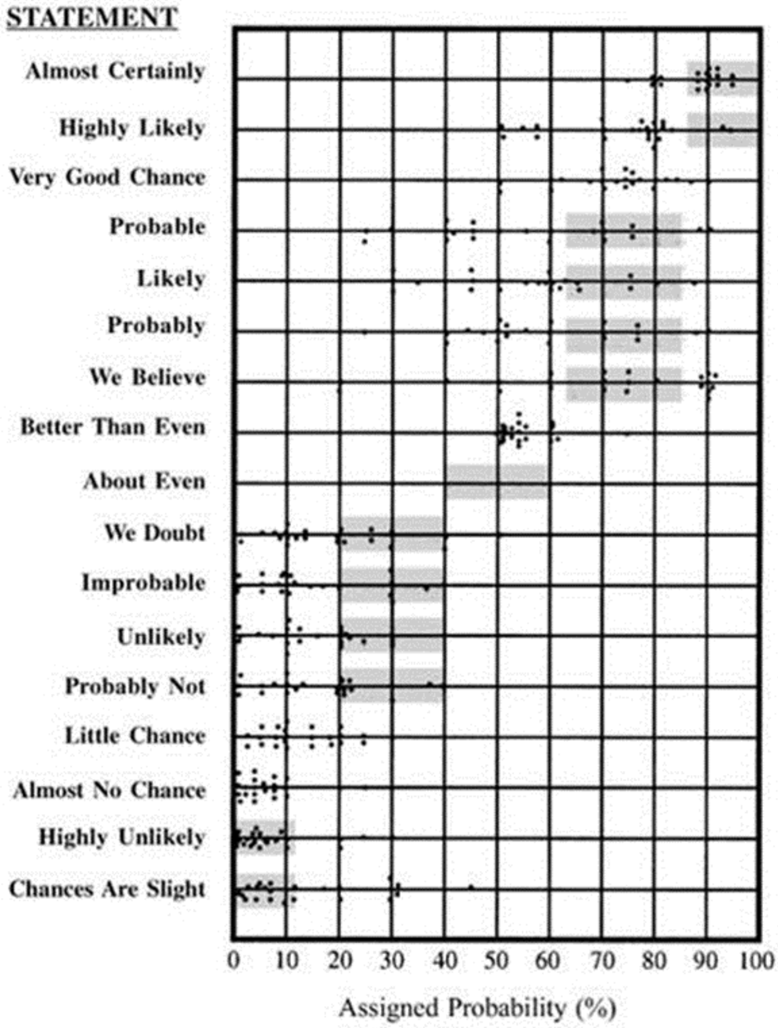


Figure 6.1: Probabilities assigned by CIA analysts to various hedges. [Jr., 1999, p. 155]

Staying within the analyst realm, Rieber [2006] requested analysts training at the Kent School (named after Sherman Kent) to assign ranges of percentages instead of specific values to a number of hedges. The results are shown in Figure 6.2.

Again one can see that the ranges of percentages range from quite narrow to relatively large, but the ranges are not necessarily identical to those in the first chart, even for identical hedges (compare *probable* in both). One can almost assume that giving the task of assigning probabilities for hedges to any random group of English-speakers will result in somewhat different

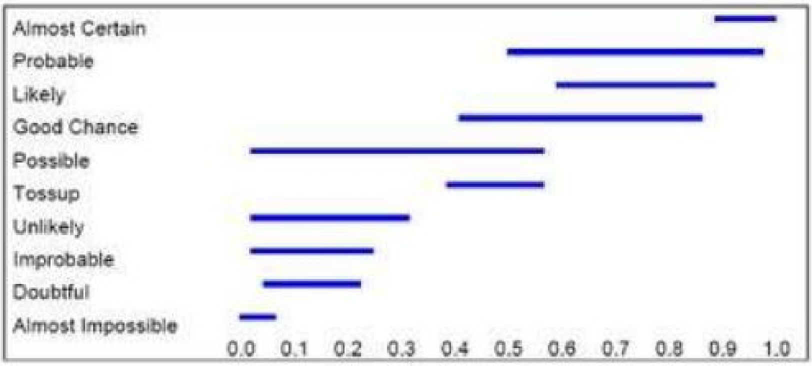


Figure 6.2: Ranges of percentages assigned to hedges by analysts in training [Rieber, 2006, p. 3]

numerical ranges.

In the decades since Kent's initial work, the US intelligence community has continued to struggle to standardize the terminology which they used to assess situations, in order to reach a common understanding of the meaning of those terms.

Ultimately, the intelligence communication settled on a standard spectrum of WEPs as shown in Figure 6.3 .

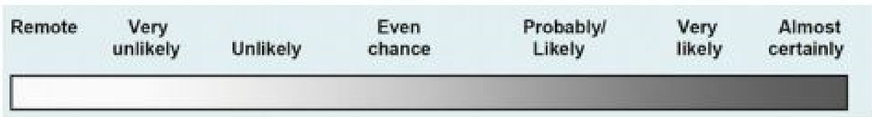


Figure 6.3: Words of Estimative Probability, as displayed in the front matter of several other recent intelligence products. via [Friedman and Zeckhauser, 2015, p. 15]

Using these "standardized" words, Wheaton [2008] had students assigning values to each of these words of estimative probability. Each student is first requested to indicate a single value for each term to represent the probability associated with that term. Then students were requested to indicate a range by identifying the lowest probability associated with each term as well as the corresponding highest probability. The results are shown in Figure 6.4 below.

In the literature from the linguistic side, there has been quite a bit of research done in which values are assigned to many expressions of uncertainty;

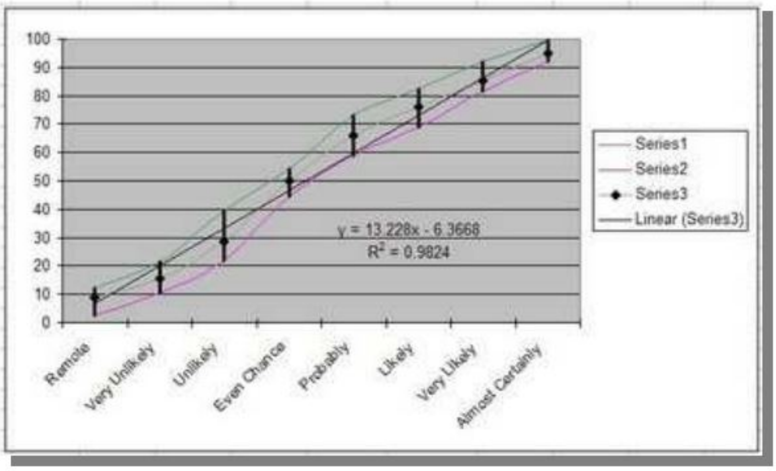


Figure 6.4: Chart based on information derived from information gathered by students. Series represent the average high score, the average low score and the average point value for each WEP, as well as an idealized trendline. [Wheaton, 2008, p. 9]

in particular modal adverbs or expressions are well represented. Indeed, there are many tables to be found in which columns of values for many that look finding numerical values which have been collected through surveys.

However, in most cases, the goal of assigning numerical values to probabilistic expressions was the means to an end, rather than the end in itself, as was the case with the examples above.

Weighting differences also appear when differing content domains are examined. The examples above were derived from the intelligence community, which actively evaluates information based upon its credibility. Examining the wider population, and, in particular, gathering information in different content domains gives insight into the consistency, or lack thereof, of the use of hedges.

Brun and Teigen [1988] investigated the numerical weights of probability expressions in three separate contexts: usage in videotaped television news reports, discussions of medical treatment effectiveness (pediatrician/patient parents conversations) and opinions on current events. Their focus was on the evaluation of not only of weighting differences between various domains (medicine, news, opinion columns), but, in the case of the medical discussion, also the differences between the understanding of the expressions between the players in the conversation (doctors, parents of the sick children). The values assigned are shown in Figures 6.5 and 6.6. The information contained in these two figures shows a strong consistency in the rankings even when different domains and contexts are involved.

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

			Treatment context					
	No context (students)		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

Figure 6.5: Probabilites ratings in the context of medical treatment [Brun and Teigen, 1988, p. 397]

Within the context of current events discussions, Brun and Teigen carried out a much more detailed analysis, involving three separate groups with three separate tasks, the results of which are shown in Figure 6.7:

> Group I (n = 16) was presented with the long list of probability phrases and asked to state the numerical probabilities (percentage certainty) associated with each verbal expression in the list. To assess the perceived ambiguity of the phrases, the subjects were also asked to judge how well they thought others would agree with their estimates. This was done by indicating within what limits they would expect to find the estimates of 90% of a large sample of respondents from the general population engaged in the same task. Finally, they were asked to select those phrases

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.

Figure 6.6: Weights assigned to probabilistic expressions used in televised news reports. [Brun and Teigen, 1988, p. 401]

they would consider to be the "best" probability expressions.

Group II (n = 24) received the shorter list of 14 probability items along with a response sheet with empty spaces for 0, 10, 20,... 100% certainty. The subjects were asked to place each of the probability expressions next to the most appropriate number.

Group III (n = 24) received the set of 14 complete probabilistic statements (context condition). The subjects first estimated the numerical probabilities (percentage certainty) associated with the underlined probability words used in the sentence, i.e., the probability intended by the source of the communication. Next, the subjects gave their own opinions on the subject matter by stating the numerical probability that they personally felt in each case was most appropriate, regardless of the probability phrase actually used in the sentence. [Brun and Teigen, 1988, p. 392]

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

	Group I			Group II		Group III (context)		
Expression	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

Figure 6.7: Results of Brun and Teigen's three-part testing numerical estimates of expression of uncertainty and perceived ambiguity [1988, p. 393]

Renooij and Witteman [1999], whose interest in the quantification of probabilistic expressions comes from the field of (Bayesian) computer modelling in medicine, evaluated three groups: medical students, other students, and the first two groups combined. Figure 6.8 below shows the resultant weightings:

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

Figure 6.8: 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 and Witteman, 1999, p. 23]

From the information recapped in Figure 6.8, they created the simplified probability scale shown in Figure 6.9.

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

Figure 6.9: Final scale with seven categories of probability expressions plus their calculated probability points. [Renooij and Witteman, 1999, p. 24]

Beylage-Haarmann [2010] takes a slightly different tack and compares the weighting of a small number of words of estimative probability by Americans and their cousins across the Atlantic. While in Figure 6.10 there are some differences, it turns out that they are not sufficiently significant to be of interest:

Bei perhaps und probably gibt es . . . größere Abweichungen zwischen den beiden getesteten Gruppen. Ebenso gehen die Einschätzungen bei certainly und maybe etwas auseinander. Die Analyse der Abweichungen ergibt jedoch keine Signifikanz für unterschiedliche Bewertungen zwischen Amerikanern und Engländern. Des Weiteren ergibt sich kein Unterschied von mehr als 10 Prozentpunkten. Es kann also davon ausgegangen werden, dass in beiden englischen Dialekten über die Bedeutung der Modalausdrücke Einigkeit herrscht.[2010, p. 76]

	Mittelwerte		n		Signifikanz 2-seitig
	USA	UK	USA	UK	
possibly	**54,73**	**55,71**	91	21	0,849
maybe	**52,31**	**59,81**	90	21	0,161
definitely	**86,69**	**87,21**	89	19	0,914
perhaps	**62,30**	**68,68**	88	19	0,179
probably	**70,76**	**79,63**	88	19	0,060
certainly	**78,48**	**69,16**	88	19	0,131

Figure 6.10: Ranking differences between native speakers from the USA and UK [Beylage-Haarmann, 2010, p. 76]

Ayyub and Klir [2006] looked at uncertainty modelling in engineering and the sciences, and presented a ranking of linguistics probabilities and translations based upon responses from studentis in that field (Figure 6.11).

One thing is eminently clear from all of the examples above: although there may be slight variations in weighting, and in spite of differences in the domains from which the test subjects were pulled, there is remarkably little variation in the ordering of items which appear on multiple lists. From this, one can quite comfortably conclude that there is a commonly accepted relative ranking of such words of estimative probability. This we will be able to exploit for our work.

Linguistic Probabilities and Translations

Rank	Phrase	No. of Responses	Mean	Median	Standard Deviation	Range
1	Highly probable	187	0.89	0.90	0.04	0.60–0.99
2	Very likely	185	0.87	0.90	0.06	0.60–0.99
3	Very probable	187	0.87	0.89	0.07	0.60–0.99
4	Quite likely	188	0.79	0.80	0.10	0.30–0.99
5	Usually	187	0.77	0.75	0.13	0.15–0.99
6	Good chance	188	0.74	0.75	0.12	0.25–0.95
7	Predictable	146	0.74	0.75	0.20	0.25–0.95
8	Likely	188	0.72	0.75	0.11	0.25–0.99
9	Probable	188	0.71	0.75	0.17	0.01–0.99
10	Rather likely	188	0.69	0.70	0.09	0.15–0.99
11	Pretty good chance	188	0.67	0.70	0.12	0.25–0.95
12	Fairly likely	188	0.66	0.70	0.12	0.15–0.95
13	Somewhat likely	187	0.59	0.60	0.18	0.20–0.92
14	Better than even	187	0.58	0.60	0.06	0.45–0.89
15	Rather	124	0.58	0.60	0.11	0.10–0.80
16	Slightly more than half the time	188	0.55	0.55	0.06	0.45–0.80
17	Slight odds in favor	187	0.55	0.55	0.08	0.05–0.75
18	Fair chance	188	0.51	0.50	0.13	0.20–0.85
19	Toss-up	188	0.50	0.50	0.00	0.45–0.52
20	Fighting chance	186	0.47	0.50	0.17	0.05–0.90
21	Slightly less than half the time	188	0.45	0.45	0.04	0.05–0.50
22	Slight odds against	185	0.45	0.45	0.11	0.10–0.99
23	Not quite even	180	0.44	0.45	0.07	0.05–0.60
24	Inconclusive	153	0.43	0.50	0.14	0.01–0.75
25	Uncertain	173	0.40	0.50	0.14	0.08–0.90
26	Possible	178	0.37	0.49	0.23	0.01–0.99
27	Somewhat unlikely	186	0.31	0.33	0.12	0.03–0.80
28	Fairly unlikely	187	0.25	0.25	0.11	0.02–0.75
29	Rather unlikely	187	0.24	0.25	0.12	0.01–0.75
30	Rather unlikely	187	0.21	0.20	0.10	0.01–0.75
31	Not very probable	187	0.20	0.20	0.12	0.01–0.60
32	Unlikely	188	0.18	0.16	0.10	0.01–0.45
33	Not much chance	186	0.16	0.15	0.09	0.01–0.45
34	Seldom	188	0.16	0.15	0.08	0.01–0.47
35	Barely possible	180	0.13	0.05	0.17	0.01–0.60
36	Faintly possible	184	0.13	0.05	0.16	0.01–0.50
37	Improbable	187	0.12	0.10	0.09	0.01–0.40
38	Quite unlikely	187	0.11	0.10	0.08	0.01–0.50
39	Very unlikely	186	0.09	0.10	0.07	0.01–0.50
40	Rare	187	0.07	0.05	0.07	0.01–0.30
41	Highly improbable	181	0.06	0.05	0.05	0.01–0.30

Source: Adapted from Lichtenstein, S. and Newman, J.R., *Psychonometric Sci.*, 9, 563–564, 1967.

Figure 6.11: Words of estimative probability in the sciences and engineering. [Ayyub and Klir, 2006, p. 154]

6.3 Hearsay, mindsay and other forms of evidentiality

Whereas the preceding section focused on "words of estimative probability", we will now begin to look at the other ways in which the writer indicates the content of the proposition may be less than certain.

Wesson and Pulford [2009] have focused on expressions conveying "mindsay" – opinion, recollection, belief – which they dubbed "expressions of confidence and doubt." Their focus for the quantification is not on identifying "universal values" but whether the context of time (past or present) affects the listener's understanding of the expression. Thus, the assignment of a numerical value to the selection of hedges is for comparative purposes only. Their results are shown in Figure 6.12.

Interestingly, they do document some differences in the weighting of expressions depending on whether those the cues are expressed in past or present tense. However, careful comparison of the values (M) which they have listed shows that, with very few exceptions, the ordering of the expressions remains identical regardless of tense.

Not all researchers have attempted to assign numerical values per se, but to examine relative strengths, in fact, assigning fuzzy values such as we saw in the preceding chapter rather than precise numbers.

Goujon [2009], who focuses on information extraction, uses Liddy et al.'s model (cf. Chapter 3) as a basis, and has assigned several linguistic forms the fuzzy values "low", "moderate" or "high" (Figure 6.13). He has also included some representative examples of both the categories and the rankings.

Not only lexical elements will affect our perception of the truth or untruth of a statement. The source from which that information is derived also plays a role. As Frajzyngier [Frajzyngier, 1985, p. 250] comments, "the 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 (in its narrowest definition, indicating the source of information, see Chapter 4 for a discussion).

$$personal\ experience < direct\ (sensory)\ evidence < hearsay \tag{6.1}$$

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

Expression	Past *M*	Past *SD*	Present *M*	Present *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. $df = 94$.

* $p < .05$, ** $p < .01$

Figure 6.12: Wesson and Pulford's weighting with focus on the effects of time (present, past) on listeners' rating of expressions of confidence and doubt. 2009, p. 154

Categories	Linguistic Forms	Dimensions and associated values
Adjectives	*douteux, incertain, peu probable*	Level: Low
	présumé, supposé,	Level: Moderate
	vraisemblable, probable, possible, envisageable, envisagé,	Level: High
Verbs	*dire, déclarer, annoncer* *penser, croire, douter, hésiter, ...*	Level: Moderate
Expressions	*selon toute vraisemblance, sans doute, à ce qu'on dit,* *il se peut que, il paraît*	Level: Moderate
Structures	*selon, d'après , de source(s)* « … »	Perspective: Reported Point of View
	si (except when followed by an adverb)	Level: Moderate
	aller + infinitive	Time: Future Time
	ne, n' (except « *n'importe* »)	Reality: Negative

Figure 6.13: Goujon's analysis of linguistic forms representing uncertainty based upon the work of Liddy et al. (cf. Chapter 3) 2009, p. 120

DeHaan [2001] proposed a cross-linguistic comparison of source evidentiality:

$$sensory < inferential < quotative \tag{6.2}$$

One could easily argue that there is an implicit weighting of the information from different types of sources in such a hierarchy. For example, while it is generally acknowledged that direct perception (e.g., "I saw") is more reliable than conveying hearsay ("he told me"), there has been no attempt to portray this difference by assigning to these expressions relative numerical values (such as we have seen with the hedges in the preceding sections) so that the values could be used to assign a numerical or fuzzy reliability weight for the information contained in the proposition, i.e., lower the weight to reflect somewhat more uncertainty or doubt.

Returning to a table we discussed in Chapter 4 (shown again here for the reader's convenience), we see that Marin-Arrese [2011] uses fuzzy designators ("high reliability," "medium validity", as well as "certainty", "probability", etc.) rather than numerical values to rank various epistemic and evidential expressions as shown in Figure 6.14.

Epistemic modality	Personal evidentiality		Mediated evidentiality	
Implicit Subjective	Explicit Subjective	Explicit/Implicit Intersubjective	Subjective Attribution: Endorsement/ Distancing	Objective Attribution: Evidential standing of Source
High Reliability: Certainty *must P* *Certainly P*	**High Validity** *I saw P* *I know P* *I say to you P*	**High Validity** *We have witnessed P* *We all know P* *Clearly P*	**High Validity: Overt/Covert Endorsement** *X has compellingly argued P* *X has revealed P*	**High Validity: Authority or High social standing of Source** *Scientists have discovered that P* *Everybody knows P*
Medium Reliability: Probability *will P* *Probably P*	**Medium Validity** *I think P*	**Medium Validity** *It seems P* *You would think P* *Apparently P*	**Medium Validity: Covert distancing** X has claimed P	**Medium Validity** *X has said P* *Reportedly P* *It is believed P*
Low Reliability: Possibility *may P* *Perhaps P*	**Low Validity** *I feel P* *I suppose P*	**Low Validity** *You get the feeling P* *Supposedly P*	**Low Validity: Overt Distancing** X has mistakenly held the view P	**Low Validity** *Some people believe P*

Figure 6.14: Fuzzy weightings for modality and evidentiality markers. [Marin-Arrese, 2011, p. 793]

Of interest are the elements under her designator "mediated evidentiality" in which she differentiates various forms of third-party information (hearsay).

It should be clear from the discussion above that the assignment of numerical values (probabilities, odds) to the lexical and grammatical elements which are of interest to us is not easy. Where it has been attempted, one can see variations in the values assigned; there are no "universally applicable" values. However, what one can clearly see is that these elements may be ordered along a scale from stronger to weaker (or higher to lower, or more true to less true, to name just a few possibilities). For example, in general, English speakers would agree to the following ordering:

$$\textit{rumor has it} < \textit{my neighbor told me} < \textit{numerous studies have shown} \tag{6.3}$$

which is reflected in Marin-Arrese's column "objective attribution": *rumor has it* falls under "low validity", *my neighbor told me* (assuming, of course, that I believe my neighbor to be at least somewhat credible) would fall under "medium validity" and the implication in *numerous studies have shown* is that the object of the discussion has been scientifically researched gives it a high validity under the argument "authority of source(s)."

As there appears to be consistency in the rankings between different groups of people surveyed on these topics. Thus we can conclude that there seems to be some sort of universal scalar for the various elements which we may exploit for our purposes.

6.4 Boosters and downtoners

Following upon the preceding observation about the universality of the ranking of various evidential elements, it should theoretically be possible to order all evidential elements on a scale. Unfortunately, natural language is very flexible, and to list all possible combinations of these elements and assign each combination a value would be difficult to say the least.

However, it turns out that there are some constructs which may assist us. For example, intensifiers may be used to weaken (downtoners) or to strengthen (amplifiers) the evidential weight of elements. That is, use of the downtoner *somewhat* weakens *likely* in *somewhat likely*, and similarly the booster *very* will turn *likely* into the stronger *very likely.* Assigning a numerical weight, say, on a scale from 0 (impossible) to 100 (fact) would result in weights in which this relation is true:

$$\textit{somewhat unlikely} < \textit{likely} < \textit{very likely} \tag{6.4}$$

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

$$\textit{very unlikely} < \textit{unlikely} < \textit{somewhat likely} \tag{6.5}$$

Basically *likely* and *unlikely* are antonymic, thus one might expect, using logic, that their negations result in identical values for the antonym, i.e., *unlikely* = *not likely*. In numerous languages including English, *unlikely* and *not likely* are in fact not two exactly equivalent expressions. They are closely related in that the general weights of the two elements in each pair (*unlikely* / *not likely* and *likely* / *not unlikely*) are nearly identical, but the negated version is somewhat "softened" in English. Indeed one could almost say that negation works as a downtoner on the "opposite", i.e., *not likely* is *likely* with a downtoner. The effect of the downtoner is minimal There are also instances of this phenomenon in Brun and Tiegen the summary in Figure 6.7 above.) The effect is mo

6.5 Conclusions

Finally, we have seen in the preceding section that rankings of various structures indicating the type and reliability of the source of information have an effect on how credible we view the information contained in the proposition, we can consider such structures to be boosters or downtoners with relation to the proposition. Thus, *rumor has it* from the relation described in the relation 6.5 above can be considered a downtoner, since it weakens the credibility of the proposition, and as a downtoner it may be assigned a value for the purposes of calculating an evidentiality score for the proposition.

In the following chapter, we will pull the pieces together to detail a model for combining various elements to derive a relative evidential scoring and demonstrate how this may be converted to numerical values, both crisp and fuzzy, for use in computer-based fusion algorithms.

Chapter 7

Putting it all together

It appears, from all this, that our eyes are uncertain. Two persons look at the same clock and there is a difference of two or three minutes in their reading of the time. One has a tendency to put back the hands, the other to advance them. Let us not too confidently try to play the part of the third person who wishes to set the first two aright; it may well happen that we are mistaken in turn. Besides, in our daily life, we have less need of certainty than of a certain approximation to certainty.

Remy de Gourmont *Philosophic Nights in Paris* [1920, p. 127]

Everything is vague to a degree you do not realize till you have tried to make it precise, and everything precise is so remote from everything that we normally think, that you cannot for a moment suppose that is what we really mean when we say what we think.... When you pass from the vague to the precise by the method of analysis and reflection that I am speaking of, you always run a certain risk of error ... you cannot very easily or simply get from these vague undeniable things to precise things which are going to retain the undeniability of the starting-point.

Bertrand Russell [2015, *Lecture 1*]

7.1 Introduction

In the preceding chapters we have examined various definitions of uncertainty in general and in natural language in particular. We have narrowed down our focus to the sentence level, specifically to the representation of uncertainty *about* the content of the sentence as opposed to the representation of uncertainty *within* the content.

In Chapter 3 we defined hedges, boosters and downtoners as used within the framework of this thesis. To reiterate, *hedges* are various lexical elements which the speaker uses to indicate the reliability of knowledge in the statement, which flag the knowledge has have been arrived at through some kind of reasoning, and which flag the knowledge as having been derived from sensory evidence or hearsay. *Boosters* intensify hedges, whereas *downtoners* weaken hedges.

The ultimate goal is to enable the assignment of a numerical (evidentiality) weight to a proposition in a sentence which represents the reliability of that proposition based upon the clues the speaker has included in the form of hedges or by other elements which convey uncertainty such as verb forms like passive voice, future tense, modal verbs, or subjunctive mood. One way to achieve this goal would be to anticipate all possible combinations of hedges, including their modification by boosters and downtoners, as well all other indicators of uncertainty, and assign each combination a value. However, this solution would be brittle, if not outright unrealistic: it is easy to miss out a potential combination. Furthermore, it is not uncommon to find that a given sentence has multiple expressions of uncertainty.

Additionally, in Chapter 6 we discussed various attempts to assign numerical values to some hedges (e.g., "words of estimative probability", and rankings to others where numerical values are not intuitive (e.g., markers of hearsay and mindsay). Our conclusion from this discussion was that there are no universal numerical values (weights) which exist for evidentials such as hedges and other markers of interest, but there appears to be a sort of universal *ordering* of these.

In this chapter, we present a methodology for flexible determination of the weight given to a proposition based upon lexical clues at the sentence level based upon the conclusions we drew from Chapter 6, including negation and a discussion of "toss-ups" and grey areas.

7.2 Polarity and the point of maximum uncertainty

In the tables we have seen in the previous chapter, words of estimative probability and other types of hedges were ranked based upon numerical values assigned by participants in research studies. Different numerical ranges were used by the different researchers, including a six-point scale and a probability scale from 0 to 100%.

In general, most readers would say that on these scales that one end of the range indicates maximum certainty (e.g., 100%) while the other indicates maximum uncertainty (e.g., 0%). However, this interpretation is erroneous in a significant way: as we approach 100%, we are indeed increasing certain that the proposition p must be true, but as we approach 0%, we are not increasingly uncertain that p is true, but rather we are ***increasingly certain that p cannot be true***. The point of maximum uncertainty lies elsewhere.

[Holmes, 1982, p. 13] captured this insight:

> 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.

A graphical representation of the relationship between these three categories is shown in Figure 7.1.

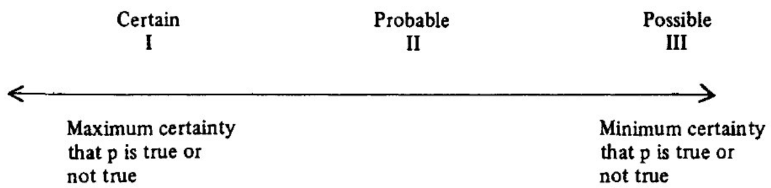


Figure 7.1: Scale of certainty which extends from maximum to minimum certainty concerning the truth or falsity of what is asserted. [Holmes, 1982, p. 13]

When Holmes' scale is opened up with ***"p is true"*** at one end of the scale and ***"p is untrue"*** at the other end, we end up with the scale shown in Figure 7.2.

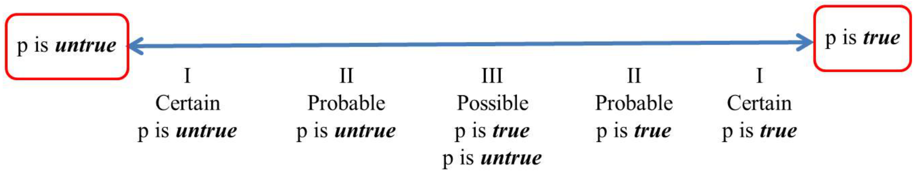


Figure 7.2: Holmes scale opened up so that ***p is true*** lies at one end of the scale and ***p is untrue*** lies at the other.

Thus, the uppermost and lowermost values of the scales represent the points of ***maximum certainty***, while minimum certainty (***maximum uncertainty***) lies in the middle of the range, as shown in Figure 7.3. The point of maximum uncertainty is the point at which we have no opinion as to the truth or untruth of p.

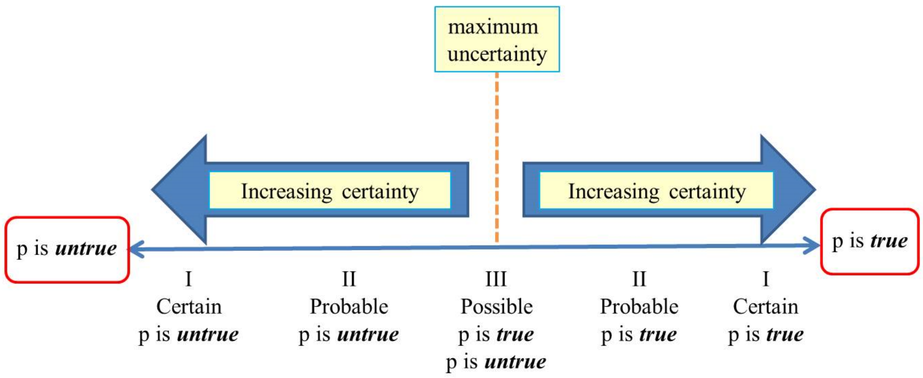


Figure 7.3: Maximum uncertainty occurs at the center of the scale, not at either end.)

Mapping some common hedges onto this scale confirms the observation that the maximum uncertainty exists in the middle of the scale, indeed as can be seen in Figure 7.4. Additionally, we can observe the effects of a booster (*very*), a downtoner (*somewhat*), as well as negation (*not*) on two hedges, namely, *likely* and *unlikely*. It is immediately apparent that the effects of the booster, downtoner and negation with relation to the scale are not identical for both hedges. The result of adding *very* to *likely* is a compound hedge which lies to the right of the original, that is, closer to ***p is true***. Adding *very* to *unlikely* results in a compound hedge which lies to the left of the original, that is, closer to ***p is untrue***. The modification of *unlikely* by the downtoner *somewhat* results in a compound which lies to the right of the unmodified hedge, i.e., closer to the point of maximum uncertainty, the rightward shift caused by the downtoner is in the same direction that the booster *very* did for *likely*. That is, the shifts caused by boosters and downtoners are dependent upon which side of the point of maximum uncertainty they lie.

Similarly, negation causes a dramatic change along the scale. If we view the point of maximum uncertainty as an axis, we can say that negation of the hedge results in a "flip" (or, the mathematically more correct term, *reflection*) around this axis, thus "changing sides" with reference to the axis. However, negation in English is not necessarily clean: while negating *likely* results in more or less the equivalent of *unlikely*, the "double negative" of *not unlikely* is not equivalent to *likely* in English. It turns out that is in addition to the reflection with respect to the axis of maximum uncertainty (from left to right), *not unlikely* lands to the left of *likely*, that is, closer to the axis of maximum of uncertainty, and thus also behaves as a downtoner.

So the problem is how to construct a method which allows for the sometimes rightward-shift, sometimes leftward-shift caused by boosters and down-

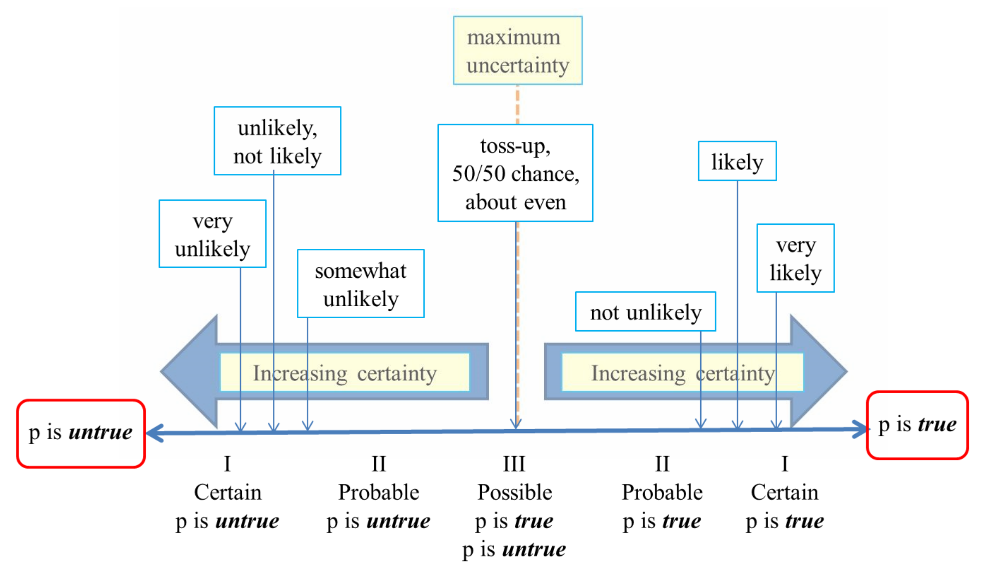


Figure 7.4: Overlaying some sample hedges onto the annotated scale.

toners, as well as for the ability to “flip” as a result of negation. It turns out that the solution is relatively simple and lies in viewing the point of maximum uncertainty as an “axis” as described above.

Since at the point of maximum uncertainty we have no opinion as to whether or not p is true or not true, we assign the value 0 to this point. From there, we assign (increasingly) positive values to those elements which indicate an increasing certainty that p is true and (increasingly) negative values to those elements which indicate an increasing certainty that p is untrue.

Among the various definitions of polarity listed in Merriam-Webster Online Dictionary [retrieved on Aug. 23, 2014] can be found “the quality or condition inherent in a body that exhibits opposite properties or powers in opposite parts or directions or that exhibits contrasted properties or powers in contrasted parts or directions.” Taking the notion of “different” or “contrasted” directions, we can say that the elements to the right of the point of maximum uncertainty have ***positive polarity***, and those to the left of that point have ***negative polarity*** as shown in Figure 7.5

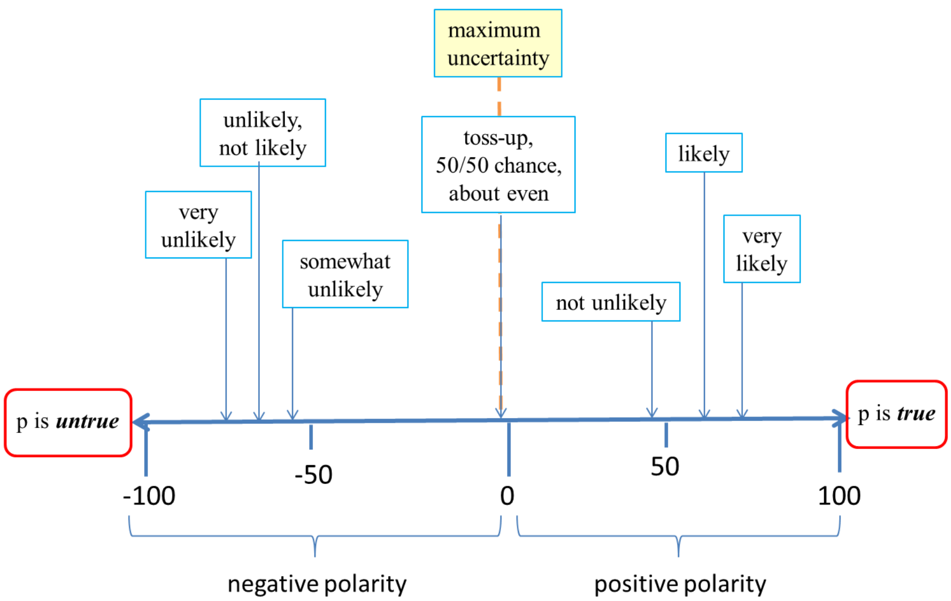


Figure 7.5: Using the point of maximum uncertainty as an axis, elements to the right are said to have positive polarity, whereas elements to the left have negative polarity.

Hedges of the first type, namely lexical elements which the speaker uses to indicate the reliability of knowledge in the statement (e.g., "words of estimative probability") can be assigned a weight. If the hedge expresses confidence that ***p is true***, i.e., lies to the right of the axis of maximum uncertainty, we call it ***positively-poled*** and assign it a positive weight between 0 and some upper limit. If the hedge expresses doubt, that is tends to assert more strongly that ***p is untrue***, we call it ***negatively-poled*** and assign it a weight between a (negative) lower limit and 0 (zero). For example, if we assign the limits 1.0 to *positive polarity* and -1.0 to *negative polarity*, we can assign *likely*, which represents confidence, a weight of 0.6, while *unlikely* is weighted at -0.6.

As will be seen in the following sections, the idea of polarity will help us to determine relative ordering of hedges, singularly and in combination, taking into account not only the effects of boosters and downtowners but also negation.

7.3 Weighting for relative ranking of "words of estimative probability"

Even though our goal at this point is to achieve a relative ranking, we will use de facto arbitrarily selected numbers on a scale (in this case the range from -1.0 to 1.0) to assist us in the process.

In Figure 7.4 several examples of hedges of the type "words of estimative probability" (WEPs) – alone, modified by boosters or downtoners, and also negated – were shown arranged along the scale from p is untrue to p is true. We will use these hedges as our initial examples.

Since we have already established in the previous chapter 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 in line with the scale shown in Figure 7.4, and in line with the polarities shown in Figure 7.5:

$$w_{likely} = 0.6 \tag{7.1}$$

$$w_{unlikely} = -0.6 \tag{7.2}$$

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: for example, *extremely* produces a bigger booster effect than *very* and *somewhat* has a very weak effect. Thus, we can assign weights to these modifiers to reflect the relative degree of modification. For example, the stronger *extremely* could be assigned an effect factor of 0.4, while *very* is assigned 0.3 to reflect its relatively weaker effect, and *somewhat* has a relatively minimal effect factor of 0.1.

We can set the generalized form for the effect of a single modifier on the original hedge to

$$effect_{modifier} = 1 - (1 - m) = m \tag{7.3}$$

where $\boldsymbol{m}$ is the weight of the modifying booster or downtoner. (It must be noted here that the expression $1 - (1 - m)$ seems to have mysteriously appeared for no apparent reason; however, while superfluous here, this expression plays a role when there are multiple modifiers present for a given hedge.) Thus the formula for the modified weight for a hedge with a single modifier is:

$$w_{modified\,hedge} = w_{original} + p * effect_{modifier} * (1 - |w_{original}|) \tag{7.4}$$

where $\boldsymbol{p}$ is the polarity of the hedge in question and $\boldsymbol{w}$ is the weight assigned to the hedge. The effect of the term $(1 - |w_{original}|)$ 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.

To demonstrate, using the value 0.3 which we assigned 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 booster $effect_{very}$ we end up with the following result:

$$\begin{aligned} w_{very\,likely} = w_{likely} + p * effect_{very} * (1 - |w_{likely}|) = \\ 0.6 + (1) * (0.3) * (1 - |0.6|) = 0.72 \end{aligned} \tag{7.5}$$

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

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

$$\begin{aligned} w_{very\,unlikely} = w_{unlikely} + p * effect_{very} * (1 - |w_{unlikely}|) = \\ -0.6 + (-1) * (0.3) * (1 - |-0.6|) = -0.72 \end{aligned} \tag{7.6}$$

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

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 value $effect_{somewhat}$ we end up with the following result:

$$w_{somewhat\,likely} = 0.6 + (1) * (-0.1) * (1 - |0.6|) = 0.56 \tag{7.7}$$

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

$$w_{somewhat\,unlikely} = -0.6 + (-1) * (-0.1) * (1 - |-0.6|) = -0.56 \tag{7.8}$$

with the result that *very unlikely* ends up to the left of *unlikely* as expected. Thus the assigned weights result in the following relation:

$$\begin{aligned} very\,unlikely < unlikely < somewhat\,unlikely < \\ somewhat\,likely < likely < very\,likely \end{aligned} \tag{7.9}$$

which are illustrated in Figure 7.6 .

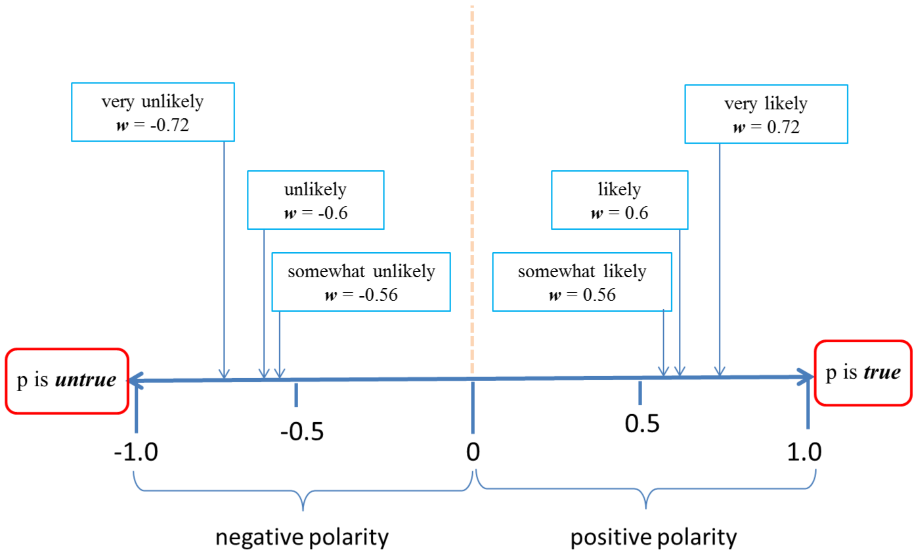


Figure 7.6: Relative weightings of *very unlikely*, *unlikely*, *somewhat unlikely*, *somewhat likely*, *likely* and *very likely* as determined by the algorithm.

It is not uncommon for humans to concatenate multiple boosters or downtoners to strengthen or weaken the hedge further. For example, it is quite common to find constructions such as *really very likely* which is even stronger than *very likely*. To account for concatenation, we need to make a modification to equation 7.3 in order to generalize the formula for concatenation of modifiers:

$$effect_{modifiers} = 1 - \prod_{i=1}^{n}(1 - m_i) \tag{7.10}$$

where m_1, m_2, ... are the boosters or downtoners which modify the hedge.

$$\begin{aligned} w_{modified\,hedge} &= w_{original} + p * (effect_{modifiers}) * (1 - |w_{original}|) \\ &= w_{original} + p * (1 - \prod_{i=1}^{n}(1 - m_i)) * (1 - |w_{original}|) \end{aligned} \tag{7.11}$$

Using this generalized equation, we can now find weights for hedges with multiple modifiers. Suppose we assign a booster weighting of 0.1 to *really*. Expanding upon the above examples of *very likely* and *very unlikely* we now calculate the values for *really very likely* and *really very unlikely*:

$$
\begin{aligned}
&w_{really\ very\ likely} = \\
&\qquad w_{likely} + p * (1 - \mathit{effect}_{really} * \mathit{effect}_{very}) * (1 - |0.6|) = \\
&\qquad 0.6 + (1)(1 - ((1 - 0.1) * (1 - 0.3)))(1 - |0.6|) = 0.748
\end{aligned} \tag{7.12}
$$

$$
\begin{aligned}
&w_{really\ very\ unlikely} = \\
&\qquad w_{unlikely} + p * (\mathit{effect}_{really} * \mathit{effect}_{very}) * (1 - |-0.6|) = \\
&\qquad -0.6 + (-1)(1 - ((1 - 0.1) * (1 - 0.3)))(1 - |-0.6|) = -0.748
\end{aligned} \tag{7.13}
$$

The resulting relative placements are shown in Figure 7.7:

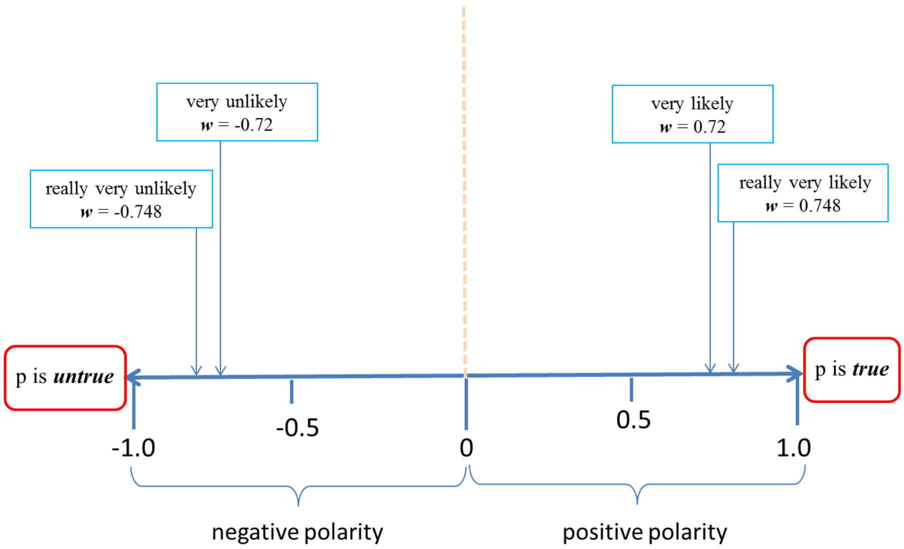


Figure 7.7: Relative weightings of *really very unlikely* and *really very likely* as determined by the algorithm.

7.4 Negation

Negation of hedges 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:

$$\mathit{effect}_{negation} = -1 \tag{7.14}$$

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

$$w_{likely} * \mathit{effect}_{negation} = 0.6 * (-1) = -0.6 \tag{7.15}$$

which is essentially the same as *unlikely*. In English, however, negation of a negatively-poled hedge generally ends up as somewhat softer than its opposite. That is, *not unlikely* is usually considered weaker than *likely* and therefore should be closer to the point of highest uncertainty. Thus, we can say that, in the case of the negation of a negatively-poled modifier, the behavior is that of is the sign change ("flip") plus a downtoner.

Thus we can differentiate the two:

$$\mathit{effect}_{negation} = \begin{cases} -1 & \text{if hedge is positively poled} \\ -1 + w_{downtowner_{negation}} & \text{if hedge is negatively poled} \end{cases} \tag{7.16}$$

If we assign the weight of the downtoner associated with negation to 0.2, the negation of *unlikely* results in

$$w_{unlikely} * \mathit{effect}_{negation} = -0.6 * (-1 + 0.2) = 0.48 \tag{7.17}$$

thus placing *not unlikely* to the left of *likely* and closer to the point of maximum uncertainty, indicating its relative weakness, as can be seen in Figure 7.8.

In order to accommodate for negation of a hedge, we must modify the equation shown in (7.17) as follows:

$$w_{negated\,hedge} = w_{hedge} * \mathit{effect}_{negation} \tag{7.18}$$

The reader will note that we have demonstrated that the negation shown in 7.18 applies to an unmodified hedge, such as *likely* or *impossible*. There is a second instance of negation, namely that of modified hedges, such as *very likely* or *quite impossible*. However, in the case of negation of a modified hedge, the negation is often applied to the modifier, rather than the hedge itself. An instance of this is *not very likely*. Whereas *not very likely* results in an equivalent expression to *unlikely*the antonym of *likely*, the modified hedge *not very likely* does not result in the antonymic *very unlikely*, but rather results in an expression with the same polarity as *very likely*, but

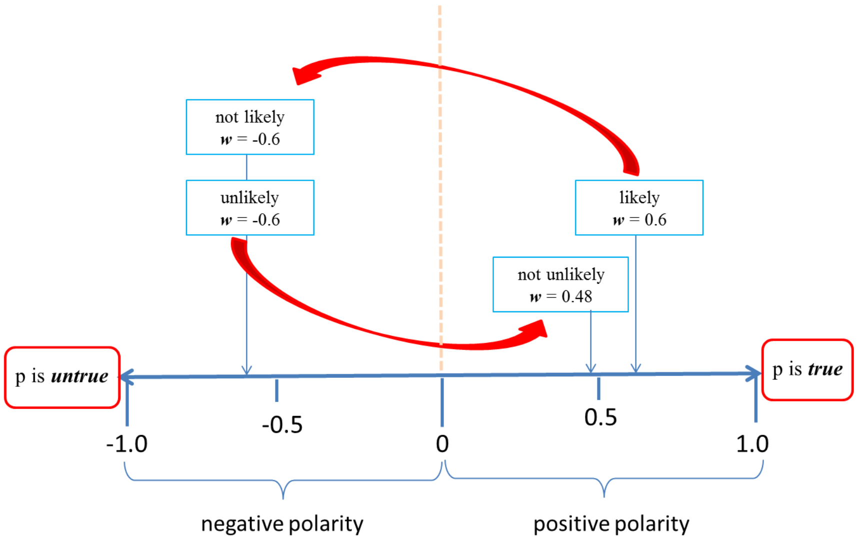


Figure 7.8: Relative weightings of *unlikely*, *not likely*, *not unlikely* and *likely* as determined by the algorithm.

weaker. In other words, in this case *not* acts as a downtoner, and thus should be handled thus.

There are, however, some exceptions to muddy things further. Using as an example the hedge *a high probability*, its negation, *not a high probability* might be evaluated as a true negation, i.e., the polarity would be flipped. Handling of such special expressions is left for the implementer to decide.

Thus far we have discussed in some depth those hedges which have either a positive or negative polarity. There is, however, relatively small set of markers which indicate maximum uncertainty, i.e., the coin is still in the air. In our model, these would appear at the boundary between the negatively-poled and positively-poled would have a numerical value of zero. As we shall discuss in the following section, the equations which we have developed up to this point are not able to appropriately handle these markers; an alternative strategy is required.

7.5 The Toss-ups

There is one place where things function somewhat differently than described above: the point of maximum uncertainty. Mathematically speaking, this is not a surprise: this point is assigned the value 0, which has neither a positive nor a negative polarity. This has an effect on our calcualtions. Take, for example, the following sentence:

(22) *The chances of rain tomorrow are 50/50.*

Clearly, assigning the weight to the the expression *50/50* is trivial, namely $w_{50/50} = 0$. The problems arise when we begin to make modifications by using boosters or downtoners.

Returning to our generalized equation for determining the effect of a booster or downtoner, 7.19 (repeated here for the reader's convenience), we are reminded, again, that $\boldsymbol{p}$ 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.

$$\begin{aligned} w_{modified\,hedge} =& w_{original} + p * \prod_{i=1}^{n}(\mathit{effect}_{modifiers}) * (1 - |w_{original}|)) \\ &= w_{original} + p * (1 - \prod_{i=1}^{n}(1 - (m_i))) * (1 - |w_{original}|) \end{aligned} \tag{7.19}$$

We can immediately see the problem: since the expression *50/50* lies neither in the positively polarized nor the negatively polarized regions of our model, but rather at the (neutral) border between the two, we cannot assign a value to $\boldsymbol{p}$ other than zero. This would result in a defaulting of a modified toss-up hedge to its original value – but since $w_{original} = 0$ in the case of a toss-up, the entire equation reduces to a zero result. This would be appropriate in the case of a modified hedge such as "very much 50/50", which simply strengthens the assertion that this is a toss-up. However, there are instances in which an effect on the toss-up hedge is reflecting through a change in weight. Examples of this would be "better than 50/50" or "somewhat less than 50/50". In such instances, the implementer could consider specific solutions for individual modified toss-up hedges; this level of specificity is outside of this thesis.

And lastly, as always, negation is also clearly an exception. Taken literally, *not 50/50* should mean any value on the scale which is not 0, regardless of polarity and regardless of strength. It is doubtful that use of negation for this hedge is intended to deliver this result; it is not unusual that *not 50/50* is followed by another hedge which clarifies the intent of the speaker, as in *chances are not 50/50 but rather closer to 75%.* In such a case, one could plead that negation of the maximum-uncertainty hedges could be ignored, with the alternative hedge combinations (*better than 50/50*, *less than 50/50*, etc.) being used for evaluation. (A possible solution would be to try to identify the alternate expression within the sentence, using that expression while ignoring the *not 50/50* – however, this, is is an implementation-specific solution and thus outside of the scope of this thesis.)

7.6 Hearsay and Mindsay

As discussed in Chapter 6, English speakers generally have little problem assigning some sort of numerical weight such as, for example, a percentage to words such as *probably*, *possibly*, *doubtful*, and so on. There are other expressions in the broader definition of hedge that we are using, namely the markers of hearsay and mindsay (hedges of types 2 and 3). While there may be hesitance to assign weights to these markers, they can be broadly, as we have also seen, ranked in some sort of order, which we will be able to exploit for our purposes.

One broad statement that can be made about hearsay and mindsay markers is this: they all weaken the credibility of the information in the proposition. Consider the following partial statements:

(23) *Die Zeit reports that...*

(24) *According to News of the World...*

(25) *Sources in the White House reported that....*

(26) *Someone told me that....*

Each of these fragments indicates that the propositional information which follows has originated from a source which is not the speaker. Just as in the preceding section we will assign weights to the markers which reflect our belief about their relative weakening influence on the proposition. These weights can be considered as "discounts" in that they reflect how much we deduct from the credibility of the information in the sentence. There may be two strategies to this assignment of weights. The first of these is simply to assign a standardized weighting for any information which is judged to have come via hearsay. The second strategy is to differentiate between named sources and anonymous sources, and may include granularity for the named sources based upon background information such as expertise of the source or our experience as to the reliability of the source. (It should be noted that the credibility of various types of sources is, strictly speaking, outside the scope of this thesis, but is worth mentioning at the juncture. There is excellent work being done by other researchers such as Rogova and Bosse [2010] in the field of information quality).

Thus, in (23) we have an example of hearsay, but the original source is named (identifiable) and is a well-respected newspaper known to be a provider of reliable information. We miight assign a minimal "discount" value of 0.05 (reflecting our belief that we have relatively little doubt) to such examples of hearsay in which original source of the information is identifiable and classified as a reliable source. Similarly, in (24) we have an example of hearsay in which the original source is named (identifiable), but in this case the original source has been known to often provide dubious information, and therefore any information provided by this source should be viewed as considerably less reliable. We might assign a value of 0.30 as a result of our doubts about the credibility of the information. On the other hand, in (25) we have an example of hearsay from an unnamed source. However, although the source is anonymous, it has been identified as being associated with an organization which quite carefully controls its information flow. We assign arelatively low "discount" weight of 0.15, based on an evaluation of the background of the source (the White House), but the anonymity creates more uncertainty. As in the previous three examples, in (26) we have hearsay, but this time we have absolutely no information about the original source. The case could also be made that the vagueness of *somebody* is a further indication of uncertainty on the part of the writer. Therefore we "discount" such hearsay quite significantly, for example, at 0.4.

Knowledge of the original source of information is very important to determining the strength of our belief in the credibility of the information being conveyed. However, it requires much background information about the information sources, including expertise, knowledge, and credibility in earlier communications. The acquisition and assessment of such source background knowledge is beyond the scope of this thesis, therefore, we will simply identify that the propositon has originated from a source other than the speaker and acknowledge that it weakens the proposition generally.

In the following sentences, we present examples of mindsay, indicating that the proposition was arrived at through a mental process and is not the result of observation of an actual event.

(27) *I believe that...*

(28) *I inferred that...*

(29) *I imagine that...*

(30) *I doubt that...*

The weights assigned to each of the mindsay hedges in the above examples will vary according to implementation; however, one would expect to see that the weighting for each will reflect its relative uncertainty (based, for example, upon the ordering discussed in Chapter 6).

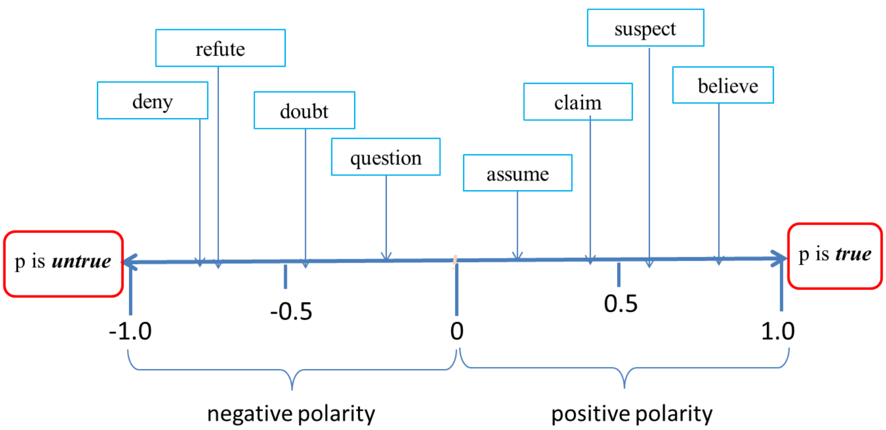


Figure 7.9: Example of relative weightings of various hearsay and mindsay markers.

Further, it should be noted that in the case of *doubt*, the mindsay marker has a negative polarity. Compare, for example, the following statements:

(31) *I believe that it will rain tomorrow.*

(32) *I doubt that it will rain tomorrow.*

The perceived likelihood of rain tomorrow (assuming I am to be believed) is higher for (31) than for (32); the difference would generally be interpreted as (31) being within the positively-poled region, while (32) falls within the negatively poled. Thus, weights assigned to markers such as *believe*, *contend*, *surmise* must reflect their positive polarity, which markers such as *doubt*, *disagree*, dispute, etc., must reflect their negative polarity. This may be seen in the examples shown in in Figure 7.9 below.

The reader will certainly note that this bears a noticeable resemblance to charts containing the hedges discussed earlier in this chapter. Likewise, similarly to hedges, hearsay and mindsay markers may be affected by boosters and downtoners:

(33) *I tend to doubt that...*

(34) *I rather suspect that...*

(35) *I very strongly believe that...*

Relative rankings of these are shown in in Figure 7.10:

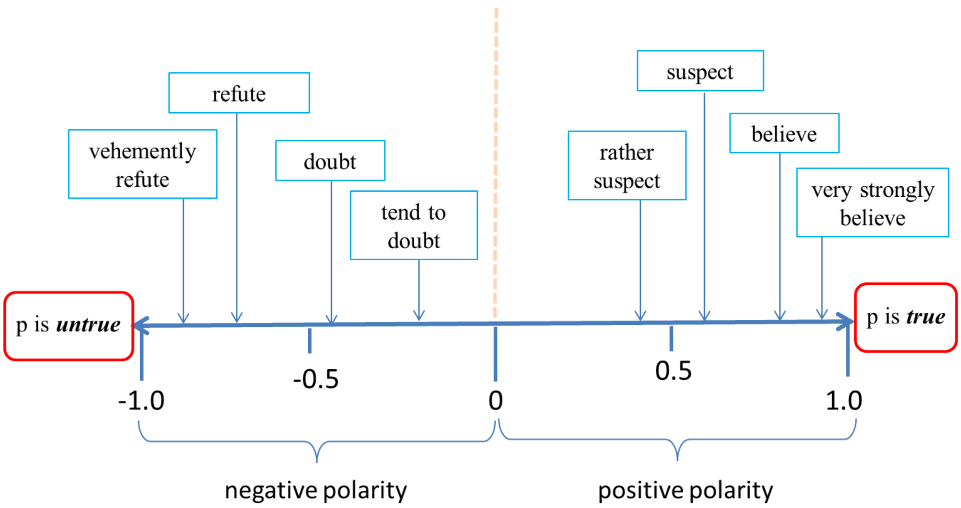


Figure 7.10: Example of relative weightings of some hearsay and mindsay markers modified by boosters and downtoners.

Negation of hearsay/mindsay markers also behaves similarly to hedges, causing a pivot around the central axis as can be seen in Figure 7.11:

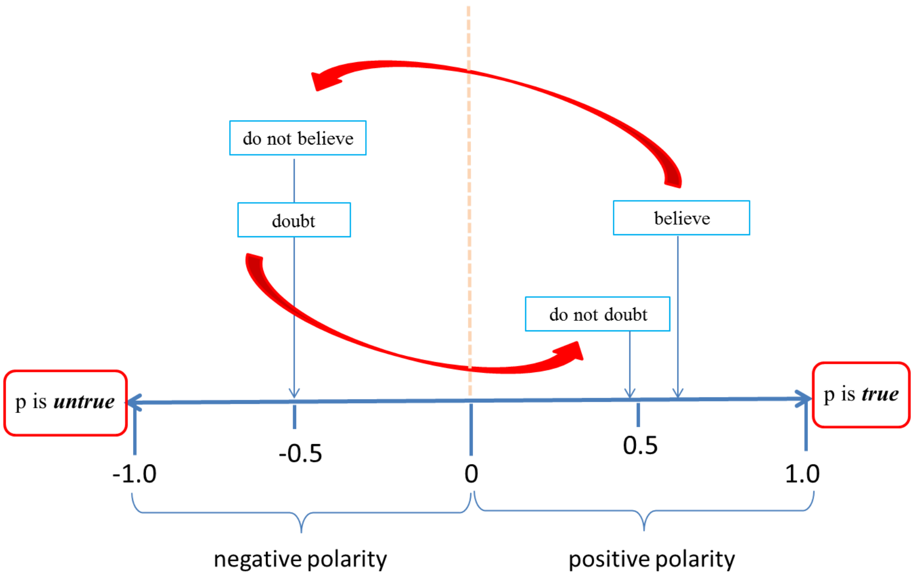


Figure 7.11: Example of relative weightings of the negation of some hearsay and mindsay markers.

In many instances of hearsay/mindsay, however, the situation is a bit more complex than presented in statements (26) through (31). All of these statements were written in the first person: *I believe*, *I doubt*, *I infer*. In many cases, there will be a combination of mindsay and hearsay, as the writer describes the beliefs, inferences and doubts of another individual or group: *researchers infer*, *sources believe*, *Mary doubts*. Thus, any formulation of mindsay which is not written in the first person needs to be weighted doubly, namely as both hearsay and mindsay, as the speaker is clearing passing on information received from another source. For example, consider the following fragments:

(36) *White House Press Secretary Josh Earnest stated he believes that. . . .*

(37) *Mary thinks it was Mark Twain who said. . . .*

In (36) the writer is reporting something which the White House Press Secretary said (*hearsay*), while Josh Earnest spoke of his belief (*mindsay*). In (37) it is even more complex: the writer repeats something Mary said (*hearsay*), Mary expressed her belief (*mindsay*) that Mark Twain (or perhaps someone else) said (*hearsay*) something quotable.

Hearsay/mindsay markers may also appear with hedges, as seen in examples (38) through (40):

(38) *I believe that it is possible that George is home now.*

(39) *My neighbor claims that it is impossible that John and Susan got married last weekend.*

(40) *We heard that there is supposed to be a big sale at the department store this weekend.*

(41) *I strongly suspect that another attack is very likely to occur within the next few weeks.*

When chained with hedges or other hearsay/mindsay markers, negatively-poled hearsay/mindsay markers functions as negation:

(42) *I doubt that it is possible that George could be home now.*

(43) *The CIA refutes the contention that the attack might have been carried out by government supporters.*

(44) *Mr. Smith vehemently denies that he may possibly run for office in the near future.*

(45) *I doubt that it is unlikely to rain this evening.*

(46) *Marjorie doesn't deny that she may be trying to have a baby.*

A simple example of interactions between hedges and negatively-poled hearsay/mindsay markers may be viewed in Figure 7.12

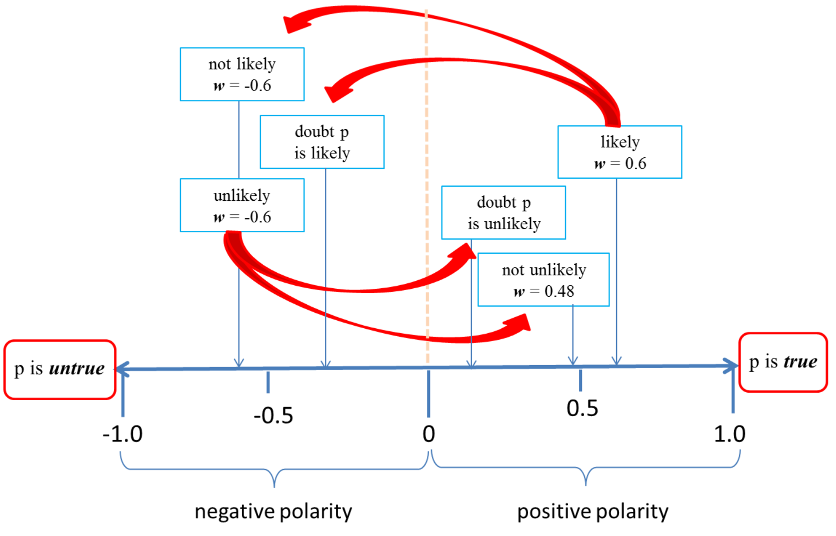


Figure 7.12: Example of relative weightings of the hedges *likely*, *not likely*, *unlikely*, and *not unlikely*, as well as *likely* and *unlikely* chained with the negatively-poled mindsay marker *doubt*.

Thus, we can see, modification of such hearsay/mindsay markers is similar to modification of hedges.

$$w_{modified\,hearsay/mindsay} = w_{hearsay/mindsay} + p * effect_{modifier} * (1 - |w_{hearsay/mindsay}|) \tag{7.20}$$

Since hearsay and mindsay always reduce the strength of the proposition, i.e., they act as downtoners, but as, in contrast to what we have seen previously, in which they softened a hedge, hearsay/mindsay markers act as *downtoners to the entire proposition*:

$$effect_{hearsay/mindsay} = (1 - |w_{hearsay/mindsay}|) \tag{7.21}$$

Chaining of hearsay/mindsay markers (*Mary thinks it was Mark Twain who said. . .*) is done multiplicatively.

As downtoners on the entire proposition, they have a direct effect on the evidentiality measure of a proposition, so we can now pull all of the pieces together and generalize overall:

$$e = \prod_{i=1}^{n} w_{hedge_i} * \prod_{j=1}^{m} w_{hearsay/mindsay_j} \tag{7.22}$$

where e = the evidentiality weight of the proposition, $hedge_i$ is the cumulative weight of all (modified or unmodified) hedges in the proposition, and $hearsay/mindsay_j$ is the cumulative effect of all hearsay/mindsay markers present.

Thus we have a simplified, but easily programmable and implemented algorithm for the calculation of evidential weights.

7.7 Grey Areas

In Chapter 6 we discussed a variety of perhaps somewhat less obvious markers of uncertainty. For example, in certain domains, the use of passive voice in English is seen as a indication of uncertainty by which the author distances herself from the assertion in the statement (and thereby signalling an unwillingness to commit fully to its veracity). In other domains, such as scholarly writing (this thesis is an immediate example thereof), the use of passive voice is quite customary and carries, in general, no particular significance (although the roots of that usage may also lie in an unwillingness to commit fully to experimental results). Therefore, the option to view the use of the passive voice as a factor in the determination of the evidential weight of a proposition will need to be determined by the implementer of this algorithm. (And, it should be noted, the treatment would be similar to that of hearsay/mindsay markers.)

In the examples which we have chosen, we have also not explicitly discussed an example such as the following:

(47) *John is a possible terrorist.*

This statement may be viewed in two ways: 1) that "John is a terrorist" is the assertion that we are evaluating and the adjective *possible* would inject some uncertainty into the veracitv of this statement; and 2) that "John is a possible terrorist" is that to which we are assigning an evidential weight, and, thus for lack of evidentials indicating uncertainty, we accept it as a true statement. The decision as to which option is most valid again lies with the implementer.

Yet another area which we have not concretely evaluated in this chapter has to do with the future tense of verbs. If the focus of one's work is predictive or proactive, for example, examining future trends in order to ward off negative events, then using information about future events or states would be of interest (and, again, treated within the algorithm similarly to hearsay/mindsay). If, however, one's goal is to develop, say, a wiki the content of which is factual, then such speculative information may be undesirable. The decision to use or ignore information which is stated using the future tense is thus, once again, up to the implementer.

The abovementioned are just a few examples of possible measures of evidential uncertainty at the sentence level which are open to interpretation; there will certainly be others, depending upon the domain which is being used. The decision to include or exclude any of these is that of the implementer.

7.8 Summary

In this chapter we have tied together various concepts from the preceding chapters: the roles of hedges, boosters and downtoners in signalling the writer's commitment to the proposition; the "univesal" ordering into a cohesive algorithm for calculating evidential weights for propositions at the sentence level which can be used by practitioners of information extraction, knowledge acquisition and information fusion to ensure the quality of information extracted from English natural language texts. Clearly, not all problems are solved, and there is still much to do with information quality in text analytics. We will look at future work in the next (and final) chapter of this thesis.

Chapter 8

Conclusions and Future Work

An approximate answer to the right problem is worth a good deal more than an exact answer to an approximate problem. **John Tukey** 1962, p. 13

In life, we make the best decisions we can with the information we have on hand. ***Agnes Kamara-umunna*** *goodreads.com [retrieved Jan. 3, 2016]*

As always with such an undertaking as a thesis like this, one hopes to contribute a small morsel to the current body of knowledge in one's field. In this case, the work should be interesting to at least two fields of research: linguistics, as we are dealing with fine points of natural language, and second, the field of application for which the results were originally intended.

For linguists, the transferability of the concepts of universal *ordering* of expressions of uncertainty, as opposed to universal values should be interesting. One might say that this applies to all natural languages in one form or another. Additionally, the mathematics underlying the interworking of hedges, boosters, downtoners as well as hearsay and mindsay markers should be, with some adaption, transferable to many, if not all, natural languages. We hope that this will prove useful.

The topic of this thesis arose as a result of the search for an implementable, pragmatic solution to determination of information quality for use in the field of information fusion. As our striving to make sense of the deluge of text-based information continues, and, in particular, as the need for understanding the reliability of that information grows, this work should help to increase the quality of the information and knowledge we acquire using automated natural language processes. This is of paramount importance for practitioners in the field.

We believe that the results of this thesis is a step in the right direction. However, we also see that there is work yet to be done, and we will discuss some aspects of this in the following sections.

8.1 Assigning weights for application

Ultimately, the purpose of this work is not an end in itself, but rather to devise a mechanism for automatically generating evidentiality weightings for simple and complex hedges that can later be used in applications using, for example, some of the mathematics for uncertainty discussed in Chapter 3. In other words, in the the area of information fusion, there are well-developed, widely-used, broadly accepted mathematical models for analyzing uncertain data based upon Bayesian probabilities, Dempster-Shafer, and fuzzy mathematics which already exist. There is no driving need within the fusion community to develop new mathematics. However, the research area of data and information fusion began several decades ago to automatically fuse data received from various sensors such as radar and sonar systems, and ground and air sensors. The data from these devices is easily represented as numerical quantities (thus "hard" information), and therefore easy to manipulate using mathematical models, and to produce algorithmic results which are assigned a numerical value representing the likelihood of the result being "true." Natural language information is less easy to quantify since it is often open to interpretation. Integration of natural language information into existing mathematical models has been problematic due to lack of adequate quantification of that information, which has remained a hurdle to the integration of both device-derived and natural information in such models.

The results of this thesis ease the quantification issue. However, the results of Chapter 7 are a bridge to the mathematics needed for fusion. In order for these to be utilized in practical applications in information fusion, they need to be converted into numerical values (in general between 0.0 and 1.0) which make sense within the mathematics of uncertainty used for this purpose.

However, this is quite an easy task, as we shall see below.

Suppose the application in which we use the propositional information we have extracted requires weights from 0.0 to 1.0, i.e., reflecting percentages. We would generated the evidential weight associated with the propositional content extracted from a sentence, locate it on our scale from -1.0 to 1.0, and then, mapping percentage values on to our scales we would end up with weights such as ***very unlikely*** being 10% or 0.1, ***unlikely*** being 30% or 0.3,***not unlikely*** being 70% (0.7), and ***very likely*** being 90% (0.9).

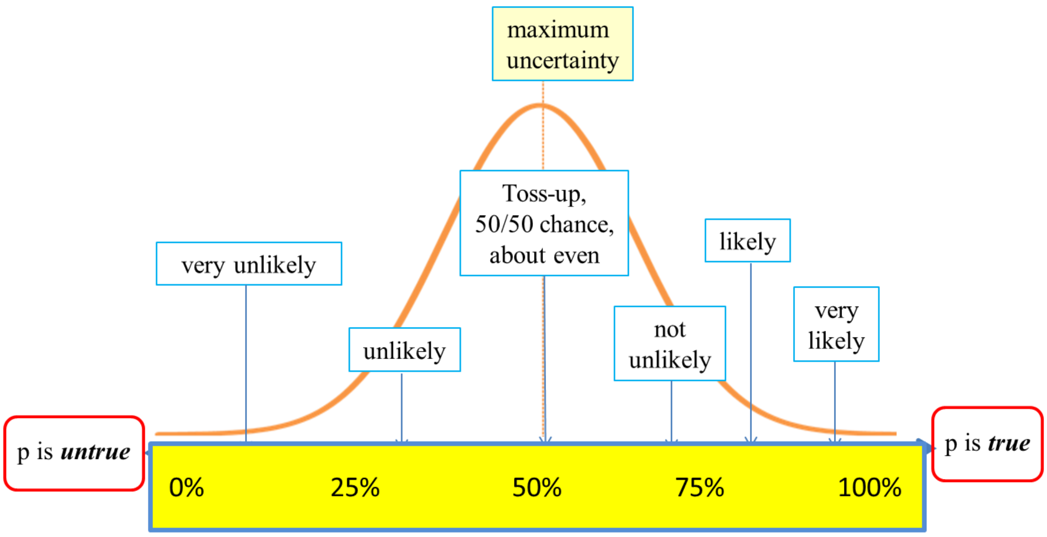


Figure 8.1: Mapping a percentage scale onto relative weightings determined by the algorithm.

Similarly, if the underlying system which implements this model relies upon fuzzy values, this mapping is also trivial. Note that the mapping of the Words of Estimative Probability from Chapter 6 very nicely coincide with the results of the algorithm.

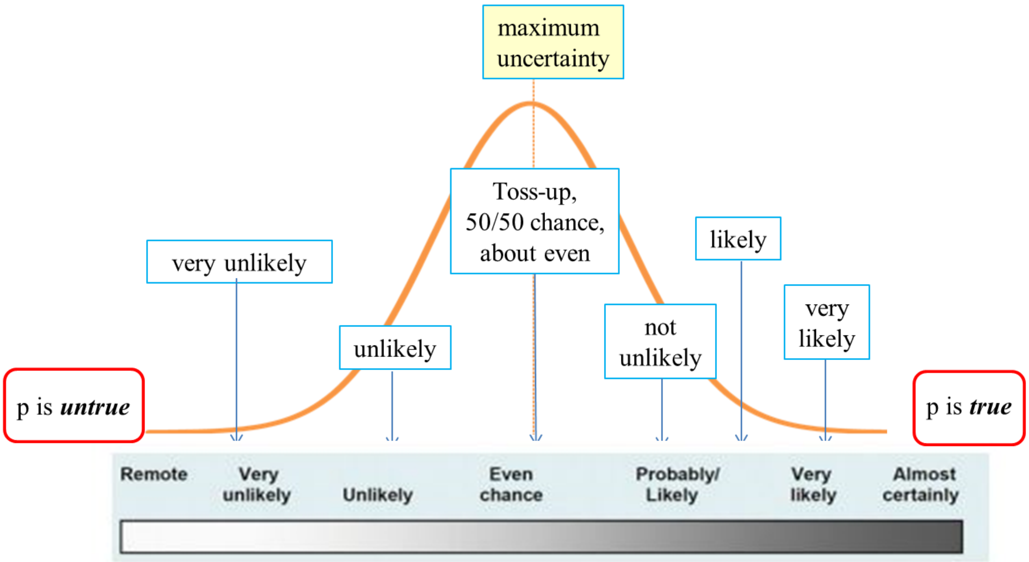


Figure 8.2: Mapping a fuzzy scale onto relative weightings determined by the algorithm.

Thus, in order to integrate the results of this thesis into another mathematical system, one needs only provide an appropriate mapping function to derive the range weights that are needed.

8.2 Open Questions

One thing which has not been discussed within this thesis has to do with the scope of the hedges and hearsay/mindsay markers within a more complex sentence – that is, we have assumed a sentence with a single proposition, with the result that all appearances of evidentiality apply to that one proposition. However, it is not always so that a sentence contains a single proposition or a single element which is hedged or re-told, therefore the practitioner is behooved to consider which sentence elements are or are not affected by these markers.

There are also open questions concerning the granularity of the hedges and modifiers. In particular, the weighting of hearsay and mindsay markers for named sources (individuals, institutions, etc.) or source types (police, government agencies, news organizations) based upon information known about the source such as area of expertise, past performance, etc. As mentioned previously in this thesis, these are important considertations but beyond the scope of the work here.

Among the areas which remain open is the ability to cross the boundaries of a given sentence to examine in the words which come before or which follow to see if there are more clues as to the veracity of the information we find in a single sentence. The work done here is limited to a single sentence (or less, in the case of compound sentences). This is a good start, but not the end by a long shot.

Bibliography

Alexandra Y. Aikhenvald. *Evidentiality.* Oxford University Press, Oxford, 2004.

Lloyd B. Anderson. Evidentials, paths of change, and mental maps: typologically regular asymmetries. In Wallace Chafe and Johanna Nichols, editors, *Evidentiality: The Linguistic Coding of Epistemology*, pages 273–312, Norwood, NJ, 1986. Ablex Publishing Corp.

Alain Auger and Jean Roy. Expression of Uncertainty in Linguistic Data. In *Proceedings of the 11th International Conference on Information Fusion*, Cologne, Germany, 2008.

Bilal M. Ayyub and George J. Klir. *Uncertainty Modeling and Analysis in Engineering and the Sciences.* JChapman and Hall/CRC, Boca Raton, 2006.

Janet Barnes. Evidentials in the Tuyuka verb. *International Journal of American Linguistics*, 50:255–271, 1984.

Bastian Beylage-Haarmann. Zwischen wahr und wahrscheinlich - Eine psycholinguistische Untersuchung der Modalsemantik im Englischen: Ranking und Unterschiede zum Deutschen. http://storage.sk.uni-bonn.de/abschlussarbeiten/magisterarbeit_haarmann.pdf, 2010.

Douglas Biber. *Variation across speech and writing.* Cambridge University Press, Cambridge, 1988.

Douglas Biber and Edward Finegan. Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. *Text*, 9:93–124, 1989.

Franz Boas. Kwakiutl grammar, with a glossary of the suffixes. *Transactions of the American Philosophical Society*, 37:203–377, 1947.

Lauren Browning. Mind-blowing growth and power of big data. *Business Insider*, 22:169–194, 2015.

Wibecke Brun and Karl Halvor Teigen. Verbal probabilities: Ambiguous, context-dependent, or both? *Organizational Behavior and Human Decision Processes*, 41(3):390 – 404, 1988. ISSN 0749-5978.

Joan Bybee and Suzanne Fleischmann. Modality in grammar and discourse: An introductory essay. In Joan Bybee and Suzanne Fleischmann, editors, *Modality in Grammar and Discourse*, pages 37–67, Amsterdam, 1995. John Benjamins.

C.A.B. Smith. Consistency in statistical inference and decision (with discussion). *Journal of the Royal Statistical Society, Series B*, 23:1–25, 1961.

Cambridge Dictionaries Online. http://dictionary.cambridge.org/ dictionary/ english /couple, retrieved on Feb. 16, 2016.

Cambridge Dictionaries Online. http://dictionary.cambridge.org /dictionary/ english/several, retrieved on Feb. 16, 2016.

Lewis Carroll. *Through the Looking-Glass.* Hayes Barton Press, Raleigh, NC, 1872.

Wallace Chafe and Johanna Nichols, editors. *Evidentiality: The linguistic coding of epistemology.* Ablex Publishing Corp., Norwood, NJ, 1986.

E.V. Clark. The principle of contrast: a constraint on language acquisition. In B. MacWhinney, editor, *Mechanisms of Language Acquisition*, Hilldale, NJ, 1987. Erlbaum.

E.V. Clark. On the logic of contrast. *Journal of Child Language*, 55:767–811, 1988.

David Clausen. HedgeHunter: A System for Hedge Detection ad Uncertainty Classification. In *Proceedings of the Fourtheenth Conference on Computational Natural Language Learning: Shared Task*, pages 120–125, Uppsala, Sweden, 2010. Assocation for Computational Linguistics.

Gudrun Clemen. *Hecken in deutschen und englischen Texten der Wirtschaftskommunikation. Eine konstrastive Analyse.*

PhD thesis, retrieved Apr. 30, 2015 from http://www.ub.uni-siegen.de/pub/diss/fb3/1999/clemen/clemen.pdf, 1998.

Peter Crompton. Hedging in Academic Writing: Some Theoretical Problems. *English for Specific Purposes*, 16:271–287, 1997.

Bruno de Finetti. *Theory of Probability, Vol. 1.* John Wiley and Sons, New York, 1974.

Remy de Gourmont. *Philosophic nights in Paris.* J.W. Luce, Boston, 1920.

Ferdinand de Haan. The place of inference within the evidential system. *International Journal of American Linguistics*, 67(2):193–219, 2001.

Ferdinand DeHaan. Evidentiality and Epistemic Modality: Setting Boundaries. *Southwest Journal of Linguistics*, (18), 1999.

Dictionary.com. http://dictionary.reference.com/browse/couple, retrieved on Feb. 16, 2016a.

Dictionary.com. http://dictionary.reference.com/browse/hedge , retrieved on Feb. 16, 2016b.

Gabriele Diewald and Elena Smirnova. *Evidentiality in German: Linguistic Realization and Regularities in Grammaticalization.* De Gruyter Mouton, Berlin, 2010.

Dider Dubois and Henri Prade. Possibility theory and its applications: a retrospective and prospective view. In *2003 IEEE International Conference on Fuzzy Systems.* IEEE, 2003.

Didier Dubois and Henri Prade. A review of fuzzy set aggregation connectives. *Annals of Mathematics and Artification Intelligence*, 32:35–66, 2001.

Didier Dubois, Henri Prade, and H.T. Nguyen. Fuzzy sets and probability: misunderstandings, bridges and gaps. In D. Dubois and H. Prade, editors, *Fundamentals of Fuzzy Sets*, pages 343–438, Boston, Mass, 2000. Kluwer.

Edgar Watson Howe. Sinner Sermons: A Selection of the Best Paragraphs of E.W. Howe, 1926.

Norman Fairclough. *Discourse and Social Change.* Blackwell, Malden, MA, 1992.

F.H. Knight, editor. *Risk, Uncertainty, and Profit.* Hart, Schaffner and Marx, Houghton Mifflin Company, Boston, 1921.

Zygmunt Frajzyngier. Truth and the indicative sentence. *Studies in Language*, 9(2):243–254, 1985.

Dick Francis. *10 lb. Penalty.* Jove Books, Berkeley Publishing Company, 1997.

Jeffrey A. Friedman and Richard Zeckhauser. Handling and Mishandling Estimative Probability: Likelihood, Confidence, and the Search for Bin Laden. *Intelligence and National Security*, 30:77–99, 2015.

Nahum Gershon. Visualization of an Imperfect World). *Computer Graphics and Applications*, 18(4):43–45, July 1998.

I.J. Good. *Probabiltiy and the weighting of evidence.* Griffin, London, 1950.

goodreads.com. http://www.goodreads.com/quotes/533915-in-life-we-make-the-best-decisions-we-can-with, retrieved Jan. 3, 2016.

Benedicte Goujon. Uncertainty Detection For Information Extraction. In *International Conference RANLP 2009*, pages 118–122, Borovets, Bulgaria, 2009.

Alan Hajek. Interpretations of Probability. *The Stanford Encyclopedia of Philosophy*, 2012.

M. Halliday. Functional diversity in language as seen from a consideration of modality and mood in English. *Foundations of Language*, 6:322–361, 1970.

Janet Holmes. Expressung Doubt and Certainty in English. *RELC Journal*, 13(9):9–28, 1982.

Ken Hyland. Hedging in academic Writing and EAP Textbooks. *English for Specific Purposes*, 13(3):239–256, 1994.

Ken Hyland. *Hedging in Scientific Research Articles.* John Benjamins, Amsterdam/Philadelphia, 1998.

Ken Hyland. Hedges, Boosters and lexical invisibility: noticing modifiers in academic texts . *Language Awareness*, 9(4):179–197, 2000.

Elly Ifantidou. *Evidentials and Relevance*. John Benjamins Publishing, 1986.

T. Inagaki. Interdependence between Safety-Control Policy and Multiple-Sensor Schemes Via Dempster-Shafer Theory. *IEEE Transactions on Reliability*, 40(2):182–188, 1991.

Roman Jakobson. *Shifters, Verbal Categories and the Russian Verb*. Harvard University, Dept of Slavic Languages and Literature, Cambridge, MA, 1957.

Anne-Laure Jousselme, Patrick Maupin, and Eloi Bosse. Uncertainty in a Situation Analysis Perspective. In *Proceedings of FUSION 2003*, Cairns, Australia, 2003. IEEE.

Richards J. Heuer Jr. *Psychology of Intelligence Analysis*. Center for the Study of Intelligence, Central Intelligence Agency, 1999.

Anne Kao and Stephen R. Poteet, editors. *Natural Language Processing and Text Mining*. Springer, London, 2007.

Sherman Kent. Words of Estimative Probability. *Studies in Intelligence*, (Fall 1964), 1964.

John Maynard Keynes, editor. *Treatise on probability*. Harper and Row, New York, 1962.

George J. Klir and Richard M. Smith. On measuring uncertainty and uncertainty-based information: Recent developments. *Annals of Mathematics and Artificial Intelligence*, 32(1):5–33, 2001.

George J. Klir and Mark J. Wierman. *Uncertainty-Based Information: Elements of Generalized Information Theory*. Studies in Fuzziness and Soft Computing. Physica-Verlag, Heidelberg, second edition, 1999.

Joerg Kohlas and Paul-Andre Monney. Representation of Evidence by Hints. In R.R.Yager, J. Kacprzyk, and M. Fedrizzi, editors, *Advances in the Dempster-Shafer Theory of Evidence*, pages 473–492. Wiley, 1994.

Joerg Kohlas and Paul-Andre Monney. A Mathematical Theory of Hints, An approach to the Dempster-Shafer Theory of Evidence. Number 425 in Lecture Notes in Economic and Mathematical Systems, Berlin, 1995. Springer Verlag.

Paul Krause and Dominic Clark. *Representing Uncertain Knowledge:,An Artificial Intelligence Approach.* Kluwer Academic Publishers, Dordrecht, 1993.

Kellyn Kruger, Ulrich Schade, and Jürgen Ziegler. Uncertainty in the fusion of information from multiple diverse sources for situation awareness . In *Proceedings of the 11thInternational Conference on Information Fusion*, pages 683–690. IEEE, 2008.

William Labov. The boundaries of words and their meanings. In C.J.N. Bailey and R. Shuy, editors, *New Ways of Analyzing Variation in English*, pages 340–373, Washington, DC, 1973. Georgetown University Press.

George Lakoff. Hedges: A study in Meaning Criteria and the Logic of Fuzzy Concepts. *Journal of Philosophical Logic*, 2:458–508, 1973.

Elizabeth D. Liddy, Noriko Kando, and Victoria L. Rubin. Certainty Categorization Model. In *The AAAI Symposium on Exploring Attitude and Affect in Text AAAI-EAAT 2004*, Stanford, CA, 2004. American Association for Artificial Intelligence.

Dennis V. Lindley. *Understanding Uncertainty.* John Wiley and Son, Hoboken, NJ, 2006.

SIL LinguaLink. What is Epistemic Modality? http://www-01.sil.org/linguistics/glossaryoflinguisticterms /WhatIsEpistemicModality.htm, retrieved Jun 10, 2015.

Juana I. Marin-Arrese. Effective vs. Epistemic Stance, and Subjectivity/Intersubjectivity in Political Discourse. A Case Study. In Anastasios Tsangalides and Roberta Facchinetti, editors, *Studies on English Modality: In Honour of Frank Palmer*, pages 23–52, Bern, 2009. Peter Lang AG.

Juana I. Marin-Arrese. Epistemic legitimizing strategies, commitment and accountability in discourse. *Discourse Studies*, 13:789–797, 2011.

Elizabeth Marshman. Expressions of uncertainty in candidate knowledge-rich contexts. A comparison in English and French specialized texts. *Terminology*, 14(1):124–147, 2008.

Merriam-Webster Dictionary. http://www.merriam-webster.com /dictionary /couple, retrieved on Feb. 16, 2016a.

Merriam-Webster Dictionary. http://www.merriam-webster.com /dictionary /some, retrieved on Feb. 16, 2016b.

Merriam-Webster Dictionary. http://www.merriam-webster.com /dictionary /hedge, retrieved on Feb. 16, 2016c.

Merriam-Webster Online Dictionary. http://www.merriam-webster.com /dictionary /polarity, retrieved on Aug. 23, 2014.

Richard Nordquist. stance (communication process). http:// grammar.about.com /od/rs /g/stanceterm.htm, retrieved Jun 13, 2015.

Janis Nuckolls and Lev Michael. Evidentials and evidential strategies in interactional and socio-cultural context. *Pragmatics and Society*, 3(2), 2014.

Oxford Dictionaries. http://www.oxforddictionaries.com /definition/english/evidence, retrieved Apr. 4, 2015.

Oxford Dictionaries. http://www.oxforddictionaries.com/definition/ english /couple, retrieved on Feb. 16, 2016a.

Oxford Dictionaries. http://www.oxforddictionaries.com/definition /english /several, retrieved on Feb. 16, 2016b.

Oxford Dictionaries. http://www.oxforddictionaries.com/definition/ english /hedge, retrieved on Feb. 16, 2016c.

Oxford English Dictionary. http://www.oed.com, retrieved on Sep. 29, 2010.

Frank R. Palmer. *Mood and Modality*. Cambridge University Press, Cambridge, 1986.

Jennifer Stevens Pappas. Qualifiers in Patient-Physician Discourse: An Analysis of Interviews from Radio Call-In Programs. *Education Linguistics*, 5 (2), 1989.

Zdzislaw Pawlak. Rough sets. *International Journal of Parallel Programming*, 11(5):341–356, 1982.

Pliny the Elder. *The Complete Works of Pliny the Elder*. Delphi Classics, 2015.

Kristin Precht. Stance moods in spoken English: Evidentiality and affect in British and American conversation. *Text*, 23:239–257, 2003.

Ellen Prince, Joel Frader, and Charles Bosk. On hedging in physician-physician discourse. In R. J. di Pietro, editor, *Linguistics and the professions*, pages 83–97, Hillsdale, NJ, 1982. Ablex.

Anna Prokofieva and Julia Hirschberg. Hedging Annotation Manual. http://www.cs.columbia.edu/ prokofieva /HedgingAnnotationManualv21.pdf, retrieved Sep 15, 2015.

Brainy Quotes. Donald Rumsfeld. http://www.brainyquote.com/quotes/ authors /d /donald_rumsfeld.html, retrieved 23 Aug 2014.

Silja Renooij and Cilia L.M. Witteman. Talking probabilities: communicating probabilistic information with words and numbers. *International Journal of Approximate Reasoning*, 22:169–194, 1999.

Steven Rieber. Communicating Uncertainty in Intelligence Analysis. http://citation.allacademic.com/meta/p100689_index.htmlf, 2006.

Roget's 21st Century Thesaurus, Third Edition. http://thesaurus.com /browse /uncertainty, retrieved on Sep. 29, 2010.

Galina L. Rogova and Eloi Bosse. Information Quality in Information Fusion. In *Proceedings of the 13th International Conference on Information Fusion*, Edinburgh, Scotland, 2010. IEEE.

Johan Rooryck. Evidentiality, Part I. *Glot International*, 5(4):125–133, April 2001a.

Johan Rooryck. Evidentiality, Part II. *Glot International*, 5(5):161–168, May 2001b.

Victoria L. Rubin. Epistemic modality: From uncertainty to certainty in the context of information seeking as interactions with texts. *Information Processing and Management*, 46(5):533 – 540, 2010. ISSN 0306-4573.

Victoria L. Rubin. Stating with Certainty or Stating with Doubt: Intercoder Reliability Results for Manual Annotation of Epistemically Modalized Statements. In *Proceedings of NAACL HLT2007, Companion Volume*, pages 141–144, Rochester, NY, April 2007. Association for Computational Linguistics.

Bertrand Russell. Am I an Atheist or an Agnostic? *The Literary Guide and Rationalist Review*, 64(7):115–116, July 1949.

Bertrand Russell. *The Philosophy of Logical Atomism*. Routledge Classics, 2015. original printing date 1918.

Michael Schrage. What Percent Is 'Slam Dunk'? Give Us Odds on Those Estimates . *Washington Post*, February 20,2005.

Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, 1976.

John Skelton. The Care and Maintenance of Hedges. *ELT Journal*, 42(1): 37–43, 1968.

Philippe Smets. Imperfect information : Imprecision - Uncertainty. Technical report, IRIDIA, Universite Libre de Bruxelles, July 2, 1999.

Philippe Smets and Robert Kennes. The Transferable Belief Model. Technical report, IRIDIA, Universite Libre de Bruxelles, 1994.

Michael Smithson. *Ignorance and Uncertainty: Emerging Paradigms*. Studies in Fuzziness and Soft Computing. Springer Verlag, New York, 1989.

Michael Smithson. Ignorance and Science. Dilemmas, Perspectives, and Prospects. *Science Communication*, 15(2):133–156, 1993.

Susan Gunelius. The data explosion in 2014 minute by minute - infographic. http://aci.info/2014/07/12/the-data-explosion-in-2014-minute-by-minute-infographic/, retrieved July 12, 2014.

Christof Tannert, Horst-Dietrich Elvers, and Burkhard Jandrig. The ethics of uncertainty. In the light of possible dangers, research becomes a moral duty. *EMBO Reports*, 8(10):892–896, 2007.

The American Heritage Dictionary of the English Language, Fourth Edition. definition of evidence. http://dictionary.reference.com/browse/evidence, retrieved on Aug. 23, 2014.

The American Heritage Dictionary of the English Language, Fourth Edition. http://dictionary.reference.com/browse/uncertainty, retrieved on Feb. 10, 2013.

Judi Thomson, Beth Hetzler, Alan MacEachren, Mark Gahegan, and Misha Pavel. A typology for visualizing uncertainty. In *Proceedings of SPIE*, volume SPIE5669, pages 146–157, 2005.

John Tukey. The future of data analysis. *Annals of Mathematical Statistics*, 33:1–67, 1962.

Irena Vassileva. Commitment and detachment in English and Bulgarian. *English for Specific Purposes*, 20:83–102, 2001.

Ignacio Vazquez and Diana Giner. Writing with Conviction: The Use of Boosters in Modelling Persuasion in Academic Discourses. *Revista Alcantina de Estudios Ingelese*, 22:219–237, 2009.

Kai von Fintel. Modality and Language. In Donald M. Borchert, editor, *Encyclopedia of Philosophy - Second Edition*, Detroit, 2012. MacMillin References USA.

David Weber. Information perspective, profile, and patterns in Quechua. In Wallace Chafe and Johanna Nichols, editors, *Evidentiality: The linguistic coding of epistemology*, page 137–155, Norwood, NJ, 1986. Ablex Publishing Corp.

Websters Online. definition of evidence. http://www.websters-online-dictionary.org/evidence, retrieved Apr. 4, 2015.

Webster's Online Dictionary. http://www.webstersonlinedictionary.org, retrieved on Sep. 29, 2010.

Caroline J. Wesson and Briony D. Pulford. Verbal Expressions of Confidence and Doubt. *Psychological Reports*, 105:151–160, 2009.

Robert R. Westmoreland. *A Semantic Theory of the English Auxiliary System.* PhD thesis, Indiana University, 1998.

Kristan J. Wheaton. What Do Words Of Estimative Probability Mean? An Exercise In Analyst Education. unpublished, 2008.

P.R.R. White. Evaluative semantics and ideological positioning in journalistic discourse: A new framework for analysis. In J. Strunck I. Lassen and T. Vestergaard, editors, *Mediating Ideology in Text and Image*, pages 37–67, Amsterdam, 2006. John Benjamins.

Thomas Willett. A Cross-Linguistic Survey of the Grammaticization of Evidentiality. *Studies in Language*, 12(1):51–97, 1988.

W.J.M. Levelt. *Speaking: From Intention to Articulation*. MIT Press, Cambridge, MA, 1989.

WordNet. definition of evidence. http://wordnetweb.princeton.edu/perl /webwn?s=evidence, retrieved Apr. 4, 2015.

Ronald R. Yager. Aggregating criteria with quantifiers. *Proceedings of the ACM SIGART International Symposium on Methodologies for Intelligent Systems, ISMIS'86*, 32(1):183–189, 1986.

Lotfi A. Zadeh. Fuzzy Sets. *Information and Control*, 8:338–358, 1965.

Lotfi A. Zadeh. Fuzzy Sets as a Basis for a Theory of Possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.

Lianwen Zhang. Uncertainty Representation Using Fuzzy Measures. In R.R.Yager, J. Kacprzyk, and M. Fedrizzi, editors, *Advances in the Dempster-Shafer theory of evidence*, pages 51–69, New York, 1994. John Wiley and Sons.

Alf C. Zimmer. A model for the interpretation of verbal predictions. *International Journal of Man-Machine Studies*, 20:121–134, 1984.