Closed Domain Question Answering using Fuzzy Semantics

Richard Bergmair, Churchill College

candidate for the degree of

M.Phil. in Computer Speech, Text, and Internet Technology

rbergmair@acm.org

July 2006

I, Richard Bergmair of Churchill College, being a candidate for the degree of M.Phil. in Computer Speech, Text, and Internet Technology, hereby declare that this dissertation and the work described in it are my own work, unaided except as specified below, and that the dissertation does not contain material that has already been used to any substantial extent for a comparable purpose.

Some text in chapters 1, 2 and in section A.1 also appears in a technical report of the Computer Laboratory (Bergmair 2006) and the conference proceedings of the 2006 IEEE World Congress on Computational Intelligence (Bergmair & Bodenhofer 2006). These publications are based on prior work carried out at the University of Derby in Austria under the supervision of Ulrich Bodenhofer. The preparation of the text-material in question and all of the experimental and software development efforts described herein were undertaken while in residence as a full-time student in Cambridge. None of the material was previously submitted in fulfillment of any degree requirements.

Ulrich Bodenhofer also read an early draft of chapter 2 and provided feedback in his capacity as an expert on fuzzy logic.

word count: approx. 14500
(the examiners are not expected to read any appendices)

Richard Bergmair
July 2, 2006

Abstract

In the present report we give a thorough exposition of our first steps towards a theory of fuzzy semantics and towards the development of a closed domain question answering system in the form of a natural language database interface that produces result sets ranked according to the degree to which they fulfill our intuitions about vague expressions in natural language, and vague adjectives in particular.

We outline our ordering-based approach to semantics and introduce some of the issues involved in modelling vagueness. We show how fuzzy sets can be used as intermediate semantic representations of vague expressions. From the family of possible fuzzy logics we pin down which one best fits our modelling needs in fuzzy semantics.

We then describe the overall design of a controlled experiment involving human subjects, the software infrastructure necessary for administrating it and the statistical analyses required to draw conclusions from the data about the adequacy of our model of fuzzy semantics. We will also discuss the results from a small-scale preliminary instantiation of this experiment.

Finally we show how we could put those theoretic insights to use in a working natural language interface to a database that produces rankings of 'small cities' or 'rainy cities near San Francisco'.

Acknowledgements

First of all I would like to thank Ann Copestake. Her prior work on RMRS and the LKB represents a major cornerstone for my own work, which is why I consider it a privilege to work under her supervision. Furthermore I would like to thank Ulrich Bodenhofer. It is thanks to his skillfull teaching and his commited though entirely voluntary mentoring that the major intricacies of multi-valued logic could become accessible to a non-mathematician like myself.

I would like to thank the University of Cambridge for providing a platform for my research, bringing together excellent resources and an inspiring environment to work in. In particular I owe thanks to the Computer Lab and Churchill College for their voluntary financial contributions to my Vancouver trip, enabling me to present some of this work at the IEEE World Congress on Computational Intelligence.

Last, but not least, my studies would never have been possible without the support I received on so many different levels from my parents, as well as the financial contributions from an EPSRC studentship and a Cambridge European Bursary. I would like to thank all benefactors who made this possible.

Finally I would like to thank my friends, colleages in the Cambridge NLIP group and subscribers of the RMRS mailing list for responding to my call for participation in the experiment.

Contents

Chapter 1

Introduction and motivation

It may well be one of the major themes of our modern industrialized society that we have been, and are continuing to be, dependent on ever increasing amounts of information to run our corporations, institutions, and daily lives. When information technology first set out to solve this problem by computerized means it promised a fountain of wisdom. What has been delivered is a flood of data. Obviously the major new issue that needs to be addressed now in almost every discipline of data processing is the evaluation of data in terms of relevance criteria such as correctness, completeness, conciseness, confidence, quantity, etc. All of these features are gradable in nature. They cannot be organized in terms of strict binary partitioning, because, by virtue of their nature, they impose weak orderings of relevance on the data they describe.

A search engine is not successful just because all of the hundreds of results it produces to a given query are truthful matches of the search expression entered. It is successful only if the first ten results it shows happen to be the most relevant. This is what PageRank enabled Google to do, and possibly the reason why they have become so successful. Operating system manufacturers have recently recognized the pressing need to provide users with similar levels of access to the floods of data accumulating on their own harddisks nowadays. Apple has already introduced its Spotlight desktop search, and Microsoft will include a similar mechanism with the upcoming system release of Vista.

From our point of view there are two major questions to be answered in the construction of such search systems: (a) How do everyday users wish to express what is relevant to them? (b) What model can a computer employ to respond to such a query in terms of an ordered result set? In our opinion, the obvious answers are as follows. (a) Users want to express what is relevant to them in the same way they express every abstraction that might be on their minds: by means of natural language. (b) A given model that infers from a query expression an unordered result set can straightforwardly be turned into a model that infers a ranked result set by moving on from bivalent logic to fuzzy logic (as we will see later on). – Therefore a model of natural language semantics based on fuzzy logic seems a promising approach towards delivering the level of interpretation of natural language necessary for applications to evaluate the gradable relevance criteria that are at the centre of attention for modern day data processing.

On the other hand, natural language is vague in nature. It has been pointed out (van Deemter 2006*a*) that seven of the ten most frequent adjectives in the BNC are vague and that children use vague adjectives among their first dozens of words. These words denote intuitive rather than strictly logical concepts and give rise to Sorites-style paradoxes. Just how many hairs does it take not to be *bald*? What income qualifies someone as being *rich*? If the denotations of all words were in fact bivalent predicates, one would often be forced to draw such a line, although every such decision boundary seems completely arbitrary from the point of view of linguistic intuition. Furthermore, if bivalent predicates were in fact the denotations of *rich* and *bald*, how could we evaluate, in a compositional fashion, whether someone is a *bald rich man* in such a way as to systematically distinguish this from a *very bald rather rich man* or a *rather bald very rich man*?

Degree modifiers like *extremely*, *very*, *quite*, *rather*, *more or less*, are highly important to natural language. Most semantic models however fail, in our opinion, to reflect the prominent role of their denotations in human intuitive reasoning. As a matter of fact these denotations turn out to be extremely difficult to work into a bivalent picture of natural language semantics. A model of language semantics that offers a proper treatment of degree modification seems to entail, almost by definition, a degree-based logic.

As a result, we believe that the study of fuzzy semantics is a promising research direction both for computational linguists seeking to model the semantics of inherently vague languages as adequately as possible and for professionals of natural language processing seeking to open up language semantics to modern applications demanding ordering based information access.

We have already presented some first steps in this direction by developing a simple framework for the syntax driven semantic analysis of context-free languages with respect to fuzzy relational semantics (Bergmair 2006). This prior work was also presented at a major conference to the fuzzy systems community (Bergmair & Bodenhofer 2006). Our general model of fuzzy semantics will also be presented in this report in chapter 2. We will follow up on this work by introducing some of the issues involved in modelling vagueness, and showing how fuzzy sets can be used as intermediate semantic representations of vague expressions. From the family of possible fuzzy logics we will then pin down that which best fits our modelling needs in fuzzy semantics.

In chapter 3 we will describe the overal design of an experiment involving human subjects, the software infrastructure necessary for administrating it and the statistical analyses required to draw conclusions from the data about the adequacy of our model of fuzzy semantics. We will also discuss the results from a small-scale preliminary instantiation of this experiment.

In chapter 4 we will show how we could put those theoretic insights to use to produce a working natural language interface to a database (NLID) that uses the fuzzy sets derived from the experiments to produce orderings of records in a database ranked according to the degree to which they fulfill our intuitions behind expressions involving vague adjectives like ‘small city’ or ‘rainy city near San Francisco’.

A major challenge in the design of the software infrastructure needed to support this work was that the results on our closed example domain might fail to generalize to different applications. This is why we took the problem to a meta-level as far as the software engineering was concerned in designing a toolkit that supports rapid prototyping of natural language database interfaces in domains characterized in a linguistic data modelling (LDM) language of our own design. Aspects of this toolkit will be introduced throughout the work as needed to support the understanding of our experimental work and our prototype NLID.

Throughout the work we will make reference to related research, especially linguistic work on vagueness as needed. For a more comprehensive treatment of the surrounding literature about mathematical approaches to fuzzy logic and psychological results on degree-based intuitive reasoning, the reader may consult appendix A.

It will get apparent throughout this report that there is quite some scope to the topic of fuzzy semantics, and we could never possibly hope to cover any substantial portion of it in a 3-month research project. But we do hope to spark some interest in this exciting topic and invite the reader to follow us, in this spirit, in making one step in this promising direction.

Chapter 2

Semantic language modelling of vagueness via fuzzy logic

In this chapter we will try to introduce the reader both to fuzzy logic and to linguistic vagueness. In particular we will show how fuzzy logic can be used to define the ordering based semantics of possibly vague linguistic concepts. By linguistic concepts we shall mean the denotations of phrases as they appear throughout phrase-structure grammars as functions of their daughter phrases. What exactly we mean by ordering based semantics and vagueness will be introduced in the next section. In the sections following that, we will consider different classes of linguistic concepts in turn, starting out with elementary predicates as the atoms of logic formulae and linguistic concepts. Then we will turn to conjunctions as those formulae arising from intersective modifications.

The goal will be to arrive at a language model that is consistent with the theory of fuzzy logic, plausible from the point of view of linguistic intuition, and successful in practical application. Here a few notes on the methodology used to arrive at such a model, and the structure of its exposition are in place.

Considering elementary predication and intersective modification in turn as linguistic phenomena, we will (1) suggest a class of plausible models from fuzzy logic and (2) draw from this class of plausible models one model by what will, at this point, be merely an educated guess in the hope that we can, at a later point, empirically establish the adequacy of the specific model choice. The first step can thus be governed by theoretic motivations, whereas an engineering approach is necessary in the second step to arrive at a practical model.

Methodologically, what we will do in the first step is to formally introduce the linguistic phenomenon by stating axioms from fuzzy logic, trying to motivate them linguistically. Here it is important to note that we will consider some but not all possible linguistic intuitions. Therefore, we make absolutely no claim to the extent that the class of plausible models established herein is consistent with all linguistic intuitions about vagueness one could possibly have.

It is important to note that fuzzy logic is really a whole family of possible multi-valued logics. As opposed to bivalent logic, which makes canonical choices for logical operators like negation, conjunction, or universal quantification, fuzzy logic offers a multitude of possible operations for each of those functions from which a choice must be made by taking into consideration the desired properties of the resulting model. We will therefore, in the second step, have to make such design choices for particular logic operations as the semantic models of particular lexical types or grammatical rules in the form of axioms that specialize on the axioms of fuzzy logic. From these we will then infer some predictions about the linguistic phenomenon under consideration, to verify whether they are consistent with our intuitions. Thereby we will show that the model resulting from our particular design is sufficient as a model for the phenomena under discussion, but we have to stress once again that we will make no claim about whether it is necessary with respect to all linguistic intuitions one could possibly have.

2.1 Ordering-based semantics

Let $X = \{x_1, x_2, x_3, ...\}$ denote the set of elements in a domain which we wish to establish a semantic language model for. That is to say, let each x_i denote a datum relevant to the application. For a search engine this could be a web-page, for an investor information system it would be a financial press release.

We can view the meaning of a natural language expression with respect to this domain as a constraint A on the data it denotes. Such a constraint can be represented by an n-ary relation $A \subseteq X^n$ on X.

We shall see that the traditional approaches to semantics, which we will introduce under the notion of partition based semantics, have $n = 1$, where we will suggest $n = 2$ in a model that we will introduce as ordering based semantics. We will make a fundamental distinction between non-gradable concepts like *mortal* and gradable concepts like *bald* and we will show that partition based semantics is in many respects inadequate as a model for the meaning of gradable concepts.

2.1.1 Non-gradable concepts and partition based semantics

Definition 1. The *partition based* semantics of a concept on domain X is a subset $A \subseteq X$, i.e. a unary relation on X.

Consider for example the non-gradable concept of mortality. Say our domain consists of humans and gods, and Socrates is in the domain. It is clear that Socrates is mortal or he is immortal, and he is not both mortal and immortal. Let T be the set of all mortals and F be the set of all immortals in the domain. In analogy to Socrates, we know that $T \cap F = \emptyset$, since nobody is both mortal and immortal and $T \cup F = X$ since everybody is mortal or immortal. Thus we can define the meaning of the non-gradable concept of mortality, as a set $A = T$. By definition $A \subseteq X$. From the set A we can now reconstruct our intuition about the concept of mortality by letting $T' = A$ and $F' = X \setminus A$. Now we know that $T' = T$, that T and F form a partition over X, and that T' and F' form a partition over X, so $F' = F$. We conclude that the partition based meaning of the non-gradable concept of mortality on the domain of all humans and gods can be written as a subset A of all humans and gods.

We call this a *partition based* approach to semantics, since it enables us to write down a partition (T, F) of the form

$$(\{x \mid x \in A\}, \{x \mid x \notin A\})$$

as an equivalent of the intuitive notion behind mortality which is

$$(\{x \mid x \text{ is mortal}\}, \{x \mid x \text{ is immortal}\})$$

At this point, traditional Aristotelian logic makes the assumption that every concept of interest to logic behaves like that of mortality, in the sense that the above argument about A and the intuitive notion of mortality equivalently holds for every concept. In fact this assumption has become so much a part of our thinking, that the above argument may seem tautologic.

We, however, will make the much weaker assumption that there is such a class of concepts and call this class of concepts non-gradable, but in the next section we will introduce gradable concepts which behave quite differently.

2.1.2 Gradable concepts and ordering based semantics

Definition 2. The *ordering based* semantics of a concept on domain X is a weak ordering on X, i.e. a binary relation $A \subseteq X \times X$ on X where for all $x, y, z \in X$

$$(x, x) \in \widetilde{A} \qquad \text{(reflexivity)} \qquad (2.1)$$

$$(x, y) \in \widetilde{A} \wedge (y, z) \in \widetilde{A} \Rightarrow (x, z) \in \widetilde{A} \qquad \text{(transitivity)} \qquad (2.2)$$

$$(x, y) \in \widetilde{A} \vee (y, x) \in \widetilde{A}, \qquad \text{(completeness)} \qquad (2.3)$$

Consider for example the gradable concept of baldness. Say our domain consists of millions of people. At least one of them has no hair at all, and at least one of them has at least 150000 hairs. Everyone in the domain who has hair, has exactly one hair more than someone else in the domain. Since baldness is a gradable concept we will find it notoriously difficult in this domain to draw the line between someone who is bald and someone who isn't, i.e. we will find it hard to provide a partition-based semantic representation for the concept of baldness, but we can still define a binary relation $\widetilde{\text{bald}}'$ in the following way.

Pick Socrates as an arbitrarily chosen element in the domain, and put (Socrates, Socrates) into A to express that Socrates is at least as bald as himself. This guarantees reflexivity. Next, for every pair of people, say Socrates and Sophocles, put (Socrates, Sophocles) into A to express that Socrates is at least as bald as Sophocles. Since it will be the case that Socrates is at least as bald as Sophocles or vice-versa, we get completeness. Now if Socrates is at least as bald as Sophocles, and Sophocles is at least as bald as Achilles, then it will also be the case that Socrates is at least as bald as Achilles, which is why we get transitivity.

We conclude that the ordering based semantics of the gradable concept of baldness on the domain of all men can be written as a weak ordering A on the set of all people. We call this an *ordering based* approach to semantics, since it enables us to write down a set of sequences of the form

$$\langle x_{i_1}, x_{i_2}, x_{i_3}, ..., x_{i_n} \rangle \text{ such that } a < b \text{ only if } (x_{i_a}, x_{i_b}) \in A$$

as an equivalent of the intuitive notion of a set of sequences of the form

$$\langle x_{i_1}, x_{i_2}, x_{i_3}, ..., x_{i_n} \rangle \text{ such that } a < b \text{ only if } x_{i_a} \text{ is at least as bald as } x_{i_b}.$$

2.1.3 Partition based semantics as a special case of ordering based semantics

Language models have to deal both with gradable and with non-gradable concepts. Therefore, in developing a theory of their semantics, we shall take special care to ensure that our ordering based model deals with non-gradable concepts as well as any traditional semantic model, while extending on the expressive power when it comes to gradable concepts.

More precisely, when we restrict the concepts under consideration to non-gradable ones we want to be able to establish an equivalence between the results obtained from our ordering based model and the results that would be obtained from a partition based one.

Definition 3. If A is the partition based semantics of a concept on domain X (i.e. $A \subseteq X$), then B is an equivalent representation in terms of ordering based semantics iff

$$(x, x) \in B \qquad \text{for all } x \in X \tag{2.4}$$

$$(x, y) \in B \wedge (y, x) \in B \qquad \text{iff } x, y \in A \vee x, y \notin A \tag{2.5}$$

$$(x, y) \in B \wedge (y, x) \notin B \qquad \text{iff } x \in A \wedge y \notin A \tag{2.6}$$

To see how this fits with intuition, turn back to the example of mortality, and let A be its partition based semantics. For each member of the domain like Socrates we can establish whether Socrates $\in A$ or Socrates $\notin A$, i.e. whether or not Socrates is mortal. How do we translate this to the semantics B of the same concept in the ordering-based semantic domain?

What we can do is to go through all elements like Socrates, Sophocles, Zeus, and Apollo in the domain. Clearly it will always be the case that $(\text{Socrates}, \text{Socrates}) \in B$ since Socrates will always be at least as mortal as himself. We will find that both Socrates $\in A$ and Sophocles $\in A$, so it will always be true that Socrates is at least as mortal as Sophocles and Sophocles is at least as mortal as Socrates so we get $(\text{Socrates}, \text{Sophocles}) \in B$ and $(\text{Sophocles}, \text{Socrates}) \in B$. Similarly if both Zeus $\notin A$ and Apollo $\notin A$, then Zeus is at least as mortal as Apollo and Apollo is at least as mortal as Zeus, so $(\text{Zeus}, \text{Apollo}) \in B$ and $(\text{Apollo}, \text{Zeus}) \in B$. On the other hand, if Socrates $\in A$ and Apollo $\notin A$, then Socrates will be at least as mortal as Apollo, but it will not be the case that Apollo is at least as mortal as Socrates, so we get $(\text{Socrates}, \text{Apollo}) \in B$ but $(\text{Apollo}, \text{Socrates}) \notin B$.

As a result we can take the partition based semantics A of the concept of mortality A, and write it as a set of partitions (T, F) of the form

$$(\{x_{i_1}, x_{i_2}, \ldots, x_{i_m}\}, \{x_{i_{m+1}}, x_{i_{m+2}}, \ldots, x_{i_n}\})$$

and then simply rewrite each as a sequence

$$\langle x_{i_1}, x_{i_2}, \ldots, x_{i_m}, x_{i_{m+1}}, x_{i_{m+2}}, \ldots, x_{i_n} \rangle,$$

fulfilling the intuition that if some x_{i_a} appears to the left of some x_{i_b}, then x_{i_a} is at least as mortal as x_{i_b}.

Theorem 1. *If A is the partition based semantics of a concept on domain X (i.e. $A \subseteq X$), and B is an equivalent representation in terms of ordering based semantics as in definition 3, then B will in fact be a weak ordering on domain X.*

2.1.4 Discussion

Degree-based semantics as traditionally considered in the linguistic literature (Cresswell 1977, Bierwisch 1989) are similar to our notion of ordering based semantics in that they introduce standards of comparison as decision boundaries that serve primarily as a means of comparing degrees of fulfillments of gradable expressions. But our approach is radically different from any traditional accounts of semantics in that it establishes denotations that are inherently relational.

In traditional accounts, the valuation of *big city* is taken to be *true* for a city x_i whose population is larger than the standard of comparison. In our account of semantics a valuation of *big city* is not possible for a single city x_i. Only when compared to another city x_j is it possible to say *big city* is *true* for the pair (x_i, x_j), meaning that x_i should be considered more of a big city than x_j.

Although we have claimed certain intuitions in justification of the theory established herein, we have offered little discussion on them. Maybe the question deserves to be revisited, whether there are any objects x_i, x_j, x_k in any domain where, for some gradable concept with ordering-based semantics B, one may claim that $(x_i, x_i) \notin B$ (in contradiction to reflexivity), that neither $(x_i, x_j) \in B$ nor $(x_j, x_i) \in B$ (in contradiction to completeness), or that $(x_i, x_k) \notin B$ where $(x_i, x_j) \in B$ and $(x_j, x_k) \in B$ (in contradiction to transitivity).

2.2 Fuzzy sets and relations

In the previous section we have introduced the distinction between gradable and non-gradable concepts, as well as ordering based semantics as a generalization of partition based semantics. In this section we will establish set theoretic representations of gradable and non-gradable concepts that serve as intermediate objects of computation on our way from a natual language expression to a weak ordering on a domain.

2.2.1 Characteristic functions

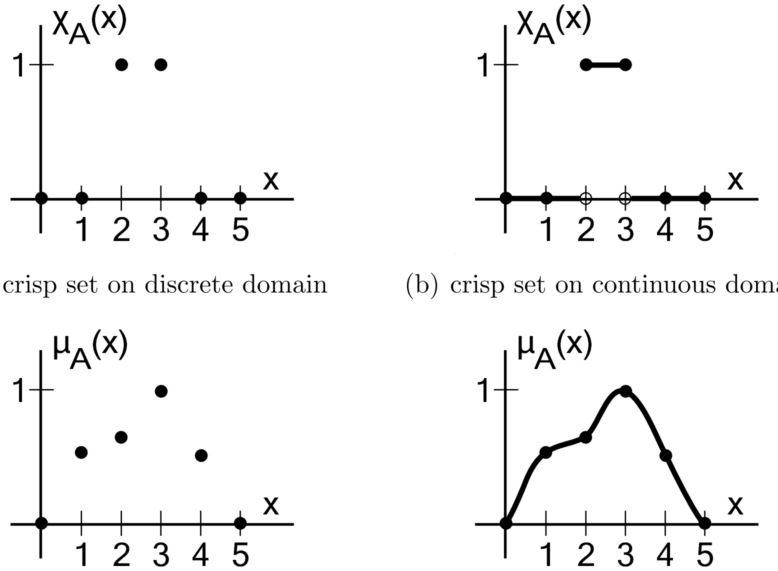
In the case of non-gradable concepts we can use the well-established definition of sets in terms of bivalent logic:

Theorem 2. *B represents the partition based semantics of a concept on domain X iff $A \in \mathcal{P}(X)$, i.e. there is a characteristic function $\chi_A : X \mapsto \{0, 1\}$, where $\chi_A(x) = 1$ iff $x \in B$ and $\chi_A(x) = 0$ iff $x \notin B$. We call A a* crisp *set.*

Fuzzy sets as proposed by Zadeh (1965) are a straightforward generalization of crisp sets over this definition. Here characteristic functions range over the whole unit interval.

Theorem 3. *B represents the ordering based semantics of a concept on domain X iff $\widetilde{A} \in \widetilde{\mathcal{P}}(X)$, i.e. there is a characteristic function $\mu_{\widetilde{A}} : X \mapsto [0, 1]$ that ranges over the whole unit interval where $\mu_{\widetilde{A}}(x) \geq \mu_{\widetilde{A}}(y)$ iff $(x, y) \in B$. We call $\widetilde{A}$ a* fuzzy *set.*

This is really the most fundamental idea of fuzzy logic: Where bivalent logic uses only two membership grades, 0 and 1, to distinguish absolute truth and absolute falseness, relevance and irrelevance, black and white, fuzzy logic allows for membership grades in



(a) crisp set on discrete domain (b) crisp set on continuous domain

(c) fuzzy set on discrete domain (d) fuzzy set on continuous domain

Figure 2.1: Characteristic functions of different kinds of sets

$]0,1[$ as well, in order to account for the shades of grey behind the notions of partial and relative truth. We call the value of $\mu_{\widetilde{A}}(x)$ for some x the degree of membership of x in $\widetilde{A}$. The tilda above is used to denote a "fuzzification", i.e. a generalization over a model based on bivalent logic towards a multi-valued logic. Figure 2.1 shows some examples of characteristic functions for fuzzy sets.

2.2.2 The standard characteristic function

How can we employ fuzzy sets and ordering-based semantics, to more adequately represent gradable concepts as expressed by elementary predications? What sort of reasoning do these representations lend themselves to?

To demonstrate this, we'll turn back to the gradable concept $\widetilde{\text{bald}}'$ and introduce the Sorites paradox as a demonstration of the limits of bivalent logic to deal with gradable concepts. If there was a crisp representation bald$'$ for the gradable concept of baldness, the following induction on bald$'$ and the number of hairs on someone's head would have to hold:

$$\begin{aligned}
\chi_{\text{bald}'}(x.\text{hair} = 0) &= 1 \\
\chi_{\text{bald}'}(x.\text{hair} = 150000) &= 0 \\
\chi_{\text{bald}'}(x.\text{hair} = h) &\Rightarrow \chi_{\text{bald}'}(x.\text{hair} = h + 1)
\end{aligned}$$

This is due to the fact that someone with no hair is bald, someone with 150000 hairs is not bald, and if someone with i hairs is bald, one additional hair won't make a difference. Obviously, by using the first statement as a basis, and the third statement as an induction, we get a contradiction with the second statement. This is why an axiom-set following this pattern is known as a Sorites *paradox*.

Using fuzzy semantics, we can easily model Sorites-like reasoning in terms of gradual degrees of fulfillment:

$$\begin{aligned}\mu_{\widetilde{\text{bald}}'}(x.\text{hair} = 0) &= 1\\ \mu_{\widetilde{\text{bald}}'}(x.\text{hair} = 1000000) &= 0\\ \mu_{\widetilde{\text{bald}}'}(x.\text{hair} = h) &\geq \mu_{\widetilde{\text{bald}}'}(h.\text{hair} = h+1)\end{aligned}$$

Obviously there is no contradiction involved here.

One often-cited critic of fuzzy logic in the linguistic community is Manfred Pinkal (Pinkal 1985, 1995). His lines of reasoning might be resolved to the following setup of the Sorites in which fuzzy logic does not solve the problem:

$$\begin{aligned}\mu_{\text{bald}'}(x.\text{hair} = 0) &= 1\\ \mu_{\text{bald}'}(x.\text{hair} = 1000000) &= 0\\ \mu_{\text{bald}'}(x.\text{hair} = h) \geq 1-\epsilon &\Rightarrow \mu_{\text{bald}'}(x.\text{hair} = h+1) \geq 1-\epsilon\end{aligned}$$

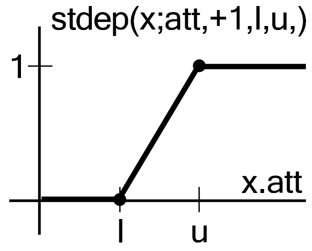
In words: *If someone with* h *hairs is bald, then so is someone with* $h+1$ *hairs.* This statement is *almost* true (i.e. ϵ is very small, but greater than zero), thus if the antecedent is almost true, so is the consequent. As a result, the paradox still holds.

Here it is important to point out that notions like *almost truth* are not formally part of fuzzy logic, although they are often used in popular expositions. In the fuzzy semantics defined herein, we take a degree of fulfillment $\mu_{\widetilde{A}}(x_1)$ of a proposition $\widetilde{A}$ about x_1 to be meaningful only when compared to the degree of fulfillment $\mu_{\widetilde{A}}(x_2)$ of $\widetilde{A}$ about some x_2. We do not employ it on a meta-level, as Pinkal does.

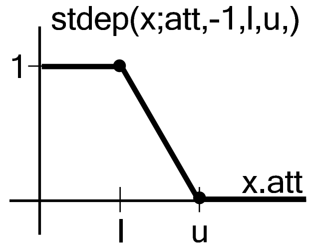
We will now move on to what we have before called the second step of our methodology for arriving at a model of fuzzy semantics by making specific design choices for our model.

Design choice: Rational valued target attributes In order to define the characteristic function for $\widetilde{\text{bald}}'$ without fixing a value for $\mu_{\widetilde{\text{bald}}'}(x)$ for every element x in the domain, we assumed in this Sorites setup that x was a complex object that is mapped to a partially ordered domain using attributes. In the context of language modelling for NLP applications, this assumption may be necessary, since it is usually not possible to obtain values of $\mu_{\widetilde{\text{bald}}'}(x)$ for every x directly, or relevance judgments about whether $(x, y) \in \widetilde{\text{bald}}'$ for every pair (x, y) as this would mean, for example, comparing every two records in a database or comparing every two web pages in a search index. A much more realistic scenario for an NLP application that deals with the concept of baldness is that some numeric measurement such as the number of a person's hairs is available as an attribute of elements in the domain. Thus, from this point on, whenever x is an element in the domain $X = \{x_1, x_2, \ldots, x_n\}$, we will use the notation $x.\text{hair}$ to refer to an attribute 'hair' associated with x. We will use an italic font and write $x.att$ whenever att is a meta-variable that ranges over all attributes available for x in the data model we are using. Futhermore we will make the simplifying assumption that this attribute is rational valued, so $x.att \in \mathbb{Q}$ for all x and att.

Hypothesis 1. A fuzzy set that is the denotation of a gradable predicate can always clearly be judged as nonincreasing or nondecreasing.



(a) $d = +1$



(b) $d = -1$

Figure 2.2: Characteristic function for standard elementary predicate.

Hypothesis 2. In a fuzzy set that is the denotation of a gradable predicate, a region can always be clearly identified in which the set does not behave like a crisp set.

Design choice: Parametric forms Given rational valued target attributes, it would be convenient to define the characteristic function $\mu_{\widetilde{\mathrm{A}}}(x)$ of a fuzzy concept as a parametric function of $x.att$ for some target attribute att. We will fix this characteristic function in the following way:

$$\mu_{\widetilde{\mathrm{A}}}(x) = \mathbf{stdep}(x; att, d, l, u) = \begin{cases} 0 & \text{if } x.att \leq l \wedge d = 1, \\ \frac{x.att - l}{u - l} & \text{if } l < x.att \leq u \wedge d = 1, \\ 1 & \text{if } u < x.att \wedge d = 1, \\ 1 & \text{if } x.att \leq l \wedge d = -1, \\ \frac{u - x.att}{u - l} & \text{if } l < x.att \leq u \wedge d = -1, \\ 0 & \text{if } u < x.att \wedge d = -1. \end{cases} \tag{2.7}$$

Figure 2.2 shows a plot of $\mathbf{stdep}(x; att, d, l, u)$ against $x.att$. This function is commonly used in fuzzy systems when parametric forms of fuzzy sets are needed. In this case the parameters are defined in the following way:

- att is the attribute of x that determines the degree of membership of x in $\widetilde{\mathrm{A}}$
- d is the decreasingness indicator, which is +1 if the degree of membership of x in $\widetilde{\mathrm{A}}$ is nondecreasing in $x.att$ and −1 if it is nonincreasing.
- l is the lower fuzziness boundary: As we set $x.att = 0$ and approach $x.att = l$, the value of l will be the maximal such value for which the concept to be defined is crisp on the domain $\{x : x.att \leq l\}$.
- u is the upper fuzziness boundary: As we set $x.att = \infty$ and approach $x.att = l$, the value of u will be the minimal such value for which the concept to be defined is crisp on the domain $\{x : x.att > u\}$.

Here, depending on d, this function is forced to be either nonincreasing or nondecreasing in the target attribute. We've seen before that $\widetilde{\mathrm{bald}}'(x)$ is indeed nonincreasing in the number of x's hairs as a result of a Sorites axiom. Another example would be $\widetilde{\mathrm{big}}'(x)$ to describe the size of a city. This will be nondecreasing in the population of x.

Furthermore, we assume that it is possible to fix two values l and u so that, for values of $x.att$ up to l and starting from u, the degree of membership of x can clearly be determined as 0 or 1. For example with $\widetilde{\text{bald}}'(x)$ there can be no dispute that someone with $l = 0$ hair is absolutely bald, and someone with $u = 150000$ hairs is not bald at all.

If we let $l = u$, we get the special case of a crisp set. Otherwise it is slightly less clear what happens between l and u. Here we made an assumption that is very common in fuzzy systems modelling, which is that we can interpolate linearly. We have to admit that this is linguistically unproven at this point, and leave the choice of the optimal interpolation as subject to future research.

Furthermore, letting

$$\mu_{\widetilde{\text{bald}}'}(x) = \mu_{\widetilde{\text{A}}}(x) = \mathbf{stdep}(x; \text{hair}, -1, 0, 150000),$$

we can still resolve the above Sorites paradox. However, the following hypothesis remains to be confirmed:

Hypothesis 3. Our assumed parametric form of fuzzy sets representing gradable predicates yields an adequate model for human intuition about gradable adjectives.

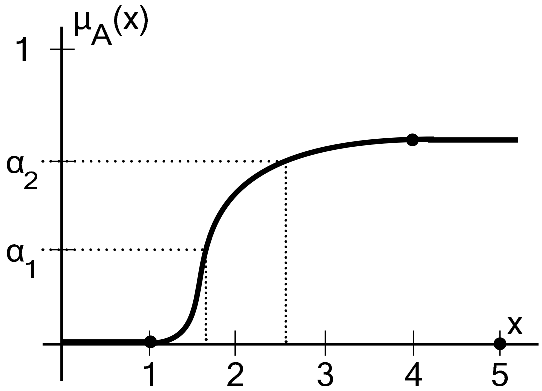
2.2.3 Discussion and future work

We have now introduced fuzzy sets as internal representations used in determining the denotations for gradable concepts. However it is important to bear in mind that this internal representation is not itself taken to be the denotation of a predicate. Rather it is the weak ordering a fuzzy set imposes on the domain that is taken to be the meaning of the predicate represented by the fuzzy set.

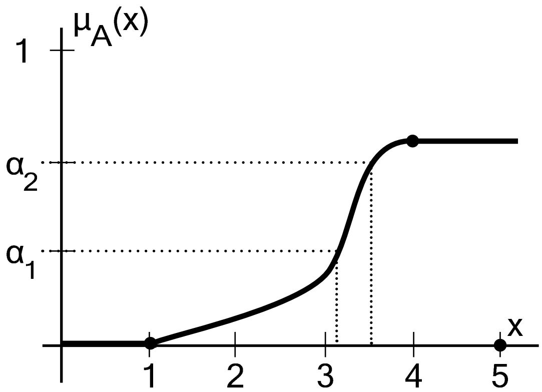
Hypothesis 4. Decision boundaries as well as fuzzy sets for different speakers may be contradictory, but each speaker is self-consistent about them.

Traditionally the theory on fuzzy logic does not make this distinction, which is highly problematic in the context of natural language. For example Moxey & Sanford (2000), and van Deemter (in personal communications) found in empirical studies that the standards of comparison used as decision boundaries for expressions like *very small* are always placed significantly different from *small* or *rather small* when observing each subject in isolation. However this effect cannot be observed when comparing different subjects. In chapter 3 we describe our own experiment which is completely consistent with these findings. It shows that some judges placed the decision boundary for *tiny* cities larger than others for *small* cities, but that no judge placed their boundary for *tiny* larger than their own boundary for *small*.

Figure 2.3 shows how this effect may be explained in the context of fuzzy semantics. Here we argue that, although fuzzy sets can be taken as an internal representation in modelling the behavior of each judge, the fuzzy sets are not universal across all judges. Two fuzzy sets for two judges may have different shapes, and naturally the distributions for the placement of decision boundaries depend on the exact shapes of these fuzzy sets when a cutoff value for the degree of fulfillment is chosen at random. Nevertheless the orderings imposed by the different fuzzy sets on their domains are much more robust, so



(a) fuzzy sets and decision boundaries



(b) different decision boundaries, same ordering

Figure 2.3: relationship between decision boundaries and orderings imposed by fuzzy sets

they might still be regarded as universal. This, of course, should be read only as a side-note at this point in support of our design choices. Empirical verification of our design must again be left as subject to future research.

Approaches in which rational valued target attributes are assumed are often objected to on the grounds that many gradable concepts use degrees that do not naturally lend themselves to measurement. The suggestion has been made that the degree of beauty of a woman could be measured in milli-Helens, the amount of beauty required to launch one ship (1/1000 of the beauty of Helen of Troy, who caused 1000 Greek warships to lift anchor). Nevertheless we argue that, in practice, the assumption of rational valued target attributes is much less of a restriction than might be assumed. If we maintain the fact that people have intuitions about the weak orderings of objects imposed by gradable concepts, it should always be possible to resort to an inverse rank, a normalized rank, a percentile or something the like.

We admit that our design choice for linear interpolation of fuzzy degrees of fulfillment may be the most objectionable we've had to make. We will return to this later, when we try to bring forward empirical evidence to find out to what extent this model can predict intuitive judgements.

2.3 Fuzzy intersections

2.3.1 Triangular norms

We can now go on to introduce the fuzzy logic operation that obtains a representation for an intersection. This is commonly done by means of triangular norms (see Klement et al. 2000).

Definition 4. Let $A, B, C \in \mathcal{P}(X)$, and

$$\chi_A(x) \ \wedge \ \chi_B(x) = \chi_C(x)$$

for all $x \in X$. Now $A \cap B = C$ iff $\wedge$ is a mapping $\wedge : \{0,1\} \times \{0,1\} \mapsto \{0,1\}$ and

$$1 \wedge 1 = 1 \tag{2.8}$$
$$1 \wedge 0 = 0 \tag{2.9}$$
$$0 \wedge 1 = 0 \tag{2.10}$$
$$0 \wedge 0 = 0 \tag{2.11}$$

The fuzzy case looks like this:

Definition 5. Let $\widetilde{A}, \widetilde{B}, \widetilde{C} \in \widetilde{\mathcal{P}}(X)$ and

$$\mu_{\widetilde{A}}(x) \ \widetilde{\wedge} \ \mu_{\widetilde{B}}(x) = \mu_{\widetilde{C}}(x)$$

for all $x \in X$. Now $\widetilde{A} \mathbin{\widetilde{\cap}} \widetilde{B} = \widetilde{C}$ iff $\widetilde{\wedge}$ is a triangular norm, i.e. a mapping $\widetilde{\wedge} : [0,1] \times [0,1] \mapsto [0,1]$ with

$$d_x \ \widetilde{\wedge} \ d_y \ = \ d_y \ \widetilde{\wedge} \ d_x \qquad \text{(commutativity)} \tag{2.12}$$
$$d_x \ \widetilde{\wedge} \ (d_y \ \widetilde{\wedge} \ d_z) \ = \ (d_x \ \widetilde{\wedge} \ d_y) \ \widetilde{\wedge} \ d_z \qquad \text{(associativity)} \tag{2.13}$$
$$d_x \leq d_y \ \Rightarrow \ d_x \ \widetilde{\wedge} \ d_z \ \leq \ d_y \ \widetilde{\wedge} \ d_z \qquad \text{(non-decreasingness)} \tag{2.14}$$
$$d_x \ \widetilde{\wedge} \ 1 \ = \ d_x \qquad \text{(neutral element)} \tag{2.15}$$

for all $d_x, d_y, d_z \in [0,1]$.

Note that, while Definition 6 gives us exactly one possible choice for a crisp conjunction operator, Definition 7 leaves open the choice for one in many possible conjunction operators. In the next section we will therefore choose a particular operator to use in our model and look at the reasoning we can do with it in our model.

2.3.2 The product t-norm

Given this notion of a fuzzy intersection, we can for example define the meaning of 'bald millionaire'. The standard approach of intersective semantics would represent this meaning as the conjunction $\text{bald}'(x) \wedge \text{millionaire}'(x)$. We have already defined $\widetilde{\text{bald}}'(x)$ as a gradable concept, and $\widetilde{\text{millionaire}}'(x)$ could be considered such a concept too. Although one might claim that there is a more technical meaning of 'millionaire', one might admit the following Sorites axioms about $\widetilde{\text{millionaire}}'(x)$:

$$\begin{aligned}
\mu_{\widetilde{\text{millionaire}}'}(x.\text{money} = 1000000) &= 1\\
\mu_{\widetilde{\text{millionaire}}'}(x.\text{money} = 0) &= 0\\
\mu_{\widetilde{\text{millionaire}}'}(x.\text{money} = m) &\geq \mu_{\widetilde{\text{millionaire}}'}(x.\text{money} = m-1)
\end{aligned}$$

By defining the concept $\widetilde{\text{bald-millionaire}}'(x)$ as

$$\mu_{\widetilde{\text{bald-millionaire}}'}(x) = \mu_{\widetilde{\text{bald}}'}(x) \mathbin{\widetilde{\wedge}} \mu_{\widetilde{\text{millionaire}}'}(x)$$

we can conclude the following about bald millionaires

$$\begin{aligned}
\mu_{\widetilde{\text{bald-millionaire}}'}(x.\text{hair} = 150000) &= 0\\
\mu_{\widetilde{\text{bald-millionaire}}'}(x.\text{money} = 0) &= 0\\
\forall m: \ \mu_{\widetilde{\text{bald-millionaire}}'}(x.\text{hair} = 0, x.\text{money} = m) &= \mu_{\widetilde{\text{millionaire}}'}(x.\text{money} = m)\\
\forall h: \ \mu_{\widetilde{\text{bald-millionaire}}'}(x.\text{hair} = h, x.\text{money} = 1000000) &= \mu_{\widetilde{\text{bald}}'}(x.\text{hair} = h)\\
\forall m: \ \mu_{\widetilde{\text{bald-millionaire}}'}(x.\text{hair} = h, x.\text{money} = m) &\geq \mu_{\widetilde{\text{bald-millionaire}}'}(x.\text{hair} = h+1, x.\text{money} = m)\\
\forall h: \ \mu_{\widetilde{\text{bald-millionaire}}'}(x.\text{hair} = h, x.\text{money} = m) &\geq \mu_{\widetilde{\text{bald-millionaire}}'}(x.\text{hair} = h, x.\text{money} = m-1)
\end{aligned}$$

only if we use a strictly increasing, rather than a general t-norm. It can be seen that, on one hand, these conclusions seem consistent with intuition and that, on the other hand, this is exactly the kind of knowledge we need in a computational procedure to produce a ranking of bald millionaires. Now the conjunction operator $\widetilde{\wedge}$ remains to be defined.

Design choice: product logic We have mentioned before that fuzzy logic is really a family of multi-valued logics. Depending on the choice for a particular triangular norm as a conjunction operator, the various bits of theory in fuzzy logic induce different logical operators for the other logical functions. There are basically four triangular norms whose induced logics are especially well understood and widely used. These are the minimum (Gödel logic), the numerical product of two rational numbers between zero and one (product logic), the Lukasiewicz t-norm (Lukasiewicz logic), and the drastic product. In our case, we have already narrowed down the family of triangular norms with properties that are theoretically desirable for the ordering based semantic reasoning we need to the family of strictly increasing triangular norms, which leaves, out of the four standard choices, only the product t-norm as it is strictly increasing on the interval $]0, 1]$. We will therefore compute the degree of truth of a conjunction by numerically multiplying the degrees of truth associated with the conjuncts. This can be done in a precise and efficient way using log-arithmetic. In the sequel we will write this operator as $\widetilde{\wedge}_*$ to remind the reader that it is implemented as a simple numeric multiplication operator $*$. However, it seems important to highlight that the underlying formalism is a logic and not a number system, so that the operator is used as a logical conjunction, not a numeric product.

2.3.3 Discussion

In this section we used a number of intuitions to support a claim that the product t-norm is best suited for modelling the fuzzy semantics of natural languages. The use of fuzzy logic, especially when using the product t-norm, is sometimes objected to on the grounds that a model based on probability theory might appear to behave identically, with probabilities offering a richer theoretic background. When we already admit that the truth of a conjunction is the product of two numbers in the unit interval, why not try to justify that the numbers are probabilities in return for the powerful and well-established theory surrounding probabilities?

Here we have to stress that a probability is the degree of expectation of an *uncertain* event x to occur, as obtained by counting the number of times this (favourable) event x has occured in the past as well as the (total) number of times any event has occurred. This degree of expectation will now be a function of the occurrences of favourable as well as unfavourable events observed. Most cases of linguistic vagueness however are not cases of such uncertainty as they do not naturally offer such closed event spaces in which to count probabilities.

For example, we have reasonable ideas of what a *mighty* oaktree is, but what do we measure here? Is it the diameter of the largest circle that can be inscribed to the shape of its trunk, or the smallest circle that can be circumscribed to it? Do the lengths of the branches matter at all? Obviously there are many problems we are facing in modelling such a vague concept, but uncertainty is not one.

Intuitively, we are not dealing with a random process in which perfect cylinders are shooting out of the earth, with the model trying to predict their diameter. A tree x may be perfectly well known, so that we can create a perfect 3d-model for x, but the problem of deciding whether x can be said to be *mighty* still remains. Say we were counting cooccurrences of events (X, R) in a semantic corpus where the cooccurence $(X = x, R = \text{mighty})$ denotes that x is referred to as mighty, and we wanted to define the degree of membership of x in $\widetilde{\text{mighty}}'$ as the probability $P(X = x|R = \text{mighty})$. If we now encounter a reference to y in our corpus, saying that y is mighty, then x will now be considered less mighty in the normalized probability space. If, on the other hand, we defined it as $P(R = \text{mighty}|X = x)$ and we encountered a reference to x, where x is referred to as healthy rather than mighty, then x would again be considered less mighty. This is certainly not what we'd expect.

The actual discrepancy here is that people have perfect intuitions about whether x

is mighty, also if they've never seen x. Probabilistically, as there is an infinite number of possible trees, any possible probability of a perfectly well-known tree x being mighty must be zero. Of course this is not about a perfectly well-known x, but rather about the similarity of x with other mighty oaktrees one has seen before. But that, of course, leads us back to the fuzzy sets we started from.

But still a more major problem with probability theory in the context of fuzzy semantics is its lack of a truth-functionality, which is, in our point of view, absolutely necessary to allow for a compositional treatment of language. The multiplication sometimes misunderstood as a probabilistic "conjunction" requires an independence assumption about its "conjuncts" which is in the context of fuzzy semantics less than easily justified. Reference is sometimes being made to a *probability logic* based on the Kolmogorov axioms in which independence assumptions are not made. Such a probability logic is in our point of view less than useful to our application, as it does not offer a closed conjunction operator.

The success of probability theory is, of course, undisputed here in its applications to syntactic language models. The scope of such models is a (stochastic) process by which syntactic elements end up in corpora. These are events in closed event-spaces that can easily be counted. We have outlined before that it is hard to define what sort of events one would exactly be looking for when modelling semantics probabilistically. But even if there was a possible definition, how could we justify any independence assumptions about semantic elements, when the fact that expressions have meaning exactly because they enter into functional relationships on the semantic level is the very basis of our endeavor to model semantics?

Using RMRS notation, the adjective phrase 'cold and rainy town' would be underspecified as $l_1 : \widetilde{\text{cold}}'(x_1), l_2 : \widetilde{\text{rainy}}'(x_2), l_3 : \widetilde{\text{town}}'(x_3)$ until a semantic model is invoked which determines that $x_1 = x_2 = x_3$ and that $l_1 = l_2 = l_3$, so we have $\widetilde{\text{cold}}'(x) \mathbin{\widetilde{\wedge}} \widetilde{\text{rainy}}'(x) \mathbin{\widetilde{\wedge}} \widetilde{\text{town}}'(x)$. It can be seen that the independence assumption is justified for the former, not for the latter, because if x is cold it may be (meteorologically) more likely to be rainy than if it was warm and if x is rainy it may be more likely to be cold than if it was dry. The effect will be that, syntactically, $l_1 : \widetilde{\text{cold}}'(x_1), l_2 : \widetilde{\text{rainy}}'(x_2)$ will be observed more often than $l_1 : \widetilde{\text{cold}}'(x_1), l_2 : \widetilde{\text{dry}}'(x_2)$ so, by conditioning syntactic probabilities appropriately, the language model may perform better.

This is due to the fact that the occurrences in the corpus will often mean $\widetilde{\text{cold}}'(x_1) \mathbin{\widetilde{\wedge}} \widetilde{\text{dry}}'(x_1)$ so, by conditioning on more lexical information, the syntactic model is actually drawing in more semantic information, which is the reason why it performs better. In order to improve the language model, one would have to draw in more and more information to condition on, until one is effectively left with a model that includes a perfect treatment of semantics by conditioning every observable event on every other event making any independence assumption impossible. The model we are left with would be a probabilistic grammar in which we are multiplying numbers between zero and one, but where no independence assumption would be made. And that, of course, is exactly what the product t-norm does for us.

2.4 Fuzzy intersections

2.4.1 Triangular norms

We can now go on to introduce the fuzzy logic operation that obtains a representation for an intersection. This is commonly done by means of triangular norms (see Klement et al. 2000).

Definition 6. Let $A, B, C \in \mathcal{P}(X)$, and

$$\chi_A(x) \ \wedge \ \chi_B(x) = \chi_C(x)$$

for all $x \in X$. Now $A \cap B = C$ iff $\wedge$ is a mapping $\wedge : \{0,1\} \times \{0,1\} \mapsto \{0,1\}$ and

$$1 \wedge 1 = 1 \quad (2.16)$$
$$1 \wedge 0 = 0 \quad (2.17)$$
$$0 \wedge 1 = 0 \quad (2.18)$$
$$0 \wedge 0 = 0 \quad (2.19)$$

The fuzzy case looks like this:

Definition 7. Let $\widetilde{A}, \widetilde{B}, \widetilde{C} \in \widetilde{\mathcal{P}}(X)$ and

$$\mu_{\widetilde{A}}(x) \ \widetilde{\wedge} \ \mu_{\widetilde{B}}(x) = \mu_{\widetilde{C}}(x)$$

for all $x \in X$. Now $\widetilde{A} \mathbin{\widetilde{\cap}} \widetilde{B} = \widetilde{C}$ iff $\widetilde{\wedge}$ is a triangular norm, i.e. a mapping $\widetilde{\wedge} : [0,1] \times [0,1] \mapsto [0,1]$ with

$$d_x \ \widetilde{\wedge} \ d_y \ = \ d_y \ \widetilde{\wedge} \ d_x \qquad \text{(commutativity)} \quad (2.20)$$
$$d_x \ \widetilde{\wedge} \ (d_y \ \widetilde{\wedge} \ d_z) \ = \ (d_x \ \widetilde{\wedge} \ d_y) \ \widetilde{\wedge} \ d_z \qquad \text{(associativity)} \quad (2.21)$$
$$d_x \leq d_y \ \Rightarrow \ d_x \ \widetilde{\wedge} \ d_z \ \leq \ d_y \ \widetilde{\wedge} \ d_z \qquad \text{(non-decreasingness)} \quad (2.22)$$
$$d_x \ \widetilde{\wedge} \ 1 \ = \ d_x \qquad \text{(neutral element)} \quad (2.23)$$

for all $d_x, d_y, d_z \in [0,1]$.

Note that, while Definition 6 gives us exactly one possible choice for a crisp conjunction operator, Definition 7 leaves open the choice for one in many possible conjunction operators. In the next section we will therefore choose a particular operator to use in our model and look at the reasoning we can do with it in our model.

2.4.2 The product t-norm

Given this notion of a fuzzy intersection, we can for example define the meaning of 'bald millionaire'. The standard approach of intersective semantics would represent this meaning as the conjunction $\text{bald}'(x) \wedge \text{millionaire}'(x)$. We have already defined $\widetilde{\text{bald}}'(x)$ as a gradable concept, and $\widetilde{\text{millionaire}}'(x)$ could be considered such a concept too. Although one might claim that there is a more technical meaning of 'millionaire', one might admit the following Sorites axioms about $\widetilde{\text{millionaire}}'(x)$:

$$\begin{aligned}
\mu_{\widetilde{\text{millionaire}}'}(x.\text{money} = 1000000) &= 1 \\
\mu_{\widetilde{\text{millionaire}}'}(x.\text{money} = 0) &= 0 \\
\mu_{\widetilde{\text{millionaire}}'}(x.\text{money} = m) &\geq \mu_{\widetilde{\text{millionaire}}'}(x.\text{money} = m - 1)
\end{aligned}$$

By defining the concept $\widetilde{\text{bald-millionaire}}'(x)$ as

$$\mu_{\widetilde{\text{bald-millionaire}}'}(x) = \mu_{\widetilde{\text{bald}}'}(x) \mathbin{\widetilde{\wedge}} \mu_{\widetilde{\text{millionaire}}'}(x)$$

we can conclude the following about bald millionaires

$$\begin{aligned}
\mu_{\widetilde{\text{bald-millionaire}}'}(x.\text{hair} = 150000) &= 0 \\
\mu_{\widetilde{\text{bald-millionaire}}'}(x.\text{money} = 0) &= 0 \\
\forall m:\ \mu_{\widetilde{\text{bald-millionaire}}'}(x.\text{hair} = 0, x.\text{money} = m) &= \mu_{\widetilde{\text{millionaire}}'}(x.\text{money} = m) \\
\forall h:\ \mu_{\widetilde{\text{bald-millionaire}}'}(x.\text{hair} = h, x.\text{money} = 1000000) &= \mu_{\widetilde{\text{bald}}'}(x.\text{hair} = h) \\
\forall m:\ \mu_{\widetilde{\text{bald-millionaire}}'}(x.\text{hair} = h, x.\text{money} = m) &\geq \mu_{\widetilde{\text{bald-millionaire}}'}(x.\text{hair} = h + 1, x.\text{money} = m) \\
\forall h:\ \mu_{\widetilde{\text{bald-millionaire}}'}(x.\text{hair} = h, x.\text{money} = m) &\geq \mu_{\widetilde{\text{bald-millionaire}}'}(x.\text{hair} = h, x.\text{money} = m - 1)
\end{aligned}$$

only if we use a strictly increasing, rather than a general t-norm. It can be seen that, on one hand, these conclusions seem consistent with intuition and that, on the other hand, this is exactly the kind of knowledge we need in a computational procedure to produce a ranking of bald millionaires. Now the conjunction operator $\widetilde{\wedge}$ remains to be defined.

Design choice: product logic We have mentioned before that fuzzy logic is really a family of multi-valued logics. Depending on the choice for a particular triangular norm as a conjunction operator, the various bits of theory in fuzzy logic induce different logical operators for the other logical functions. There are basically four triangular norms whose induced logics are especially well understood and widely used. These are the minimum (Gödel logic), the numerical product of two rational numbers between zero and one (product logic), the Lukasiewicz t-norm (Lukasiewicz logic), and the drastic product. In our case, we have already narrowed down the family of triangular norms with properties that are theoretically desirable for the ordering based semantic reasoning we need to the family of strictly increasing triangular norms, which leaves, out of the four standard choices, only the product t-norm as it is strictly increasing on the interval $]0, 1]$. We will therefore compute the degree of truth of a conjunction by numerically multiplying the degrees of truth associated with the conjuncts. This can be done in a precise and efficient way using log-arithmetic. In the sequel we will write this operator as $\widetilde{\wedge}_*$ to remind the reader that it is implemented as a simple numeric multiplication operator $*$. However, it seems important to highlight that the underlying formalism is a logic and not a number system, so that the operator is used as a logical conjunction, not a numeric product.

2.4.3 Discussion

In this section we used a number of intuitions to support a claim that the product t-norm is best suited for modelling the fuzzy semantics of natural languages. The use of fuzzy logic, especially when using the product t-norm, is sometimes objected to on the grounds that a model based on probability theory might appear to behave identically, with probabilities offering a richer theoretic background. When we already admit that the truth of a conjunction is the product of two numbers in the unit interval, why not try to justify that the numbers are probabilities in return for the powerful and well-established theory surrounding probabilities?

Here we have to stress that a probability is the degree of expectation of an *uncertain* event x to occur, as obtained by counting the number of times this (favourable) event x has occured in the past as well as the (total) number of times any event has occurred. This degree of expectation will now be a function of the occurrences of favourable as well as unfavourable events observed. Most cases of linguistic vagueness however are not cases of such uncertainty as they do not naturally offer such closed event spaces in which to count probabilities.

For example, we have reasonable ideas of what a *mighty* oaktree is, but what do we measure here? Is it the diameter of the largest circle that can be inscribed to the shape of its trunk, or the smallest circle that can be circumscribed to it? Do the lengths of the branches matter at all? Obviously there are many problems we are facing in modelling such a vague concept, but uncertainty is not one.

Intuitively, we are not dealing with a random process in which perfect cylinders are shooting out of the earth, with the model trying to predict their diameter. A tree x may be perfectly well known, so that we can create a perfect 3d-model for x, but the problem of deciding whether x can be said to be *mighty* still remains. Say we were counting cooccurrences of events (X, R) in a semantic corpus where the cooccurence $(X = x, R = \text{mighty})$ denotes that x is referred to as mighty, and we wanted to define the degree of membership of x in $\widetilde{\text{mighty}}'$ as the probability $P(X = x|R = \text{mighty})$. If we now encounter a reference to y in our corpus, saying that y is mighty, then x will now be considered less mighty in the normalized probability space. If, on the other hand, we defined it as $P(R = \text{mighty}|X = x)$ and we encountered a reference to x, where x is referred to as healthy rather than mighty, then x would again be considered less mighty. This is certainly not what we'd expect.

The actual discrepancy here is that people have perfect intuitions about whether x is mighty, also if they've never seen x. Probabilistically, as there is an infinite number of possible trees, any possible probability of a perfectly well-known tree x being mighty must be zero. Of course this is not about a perfectly well-known x, but rather about the similarity of x with other mighty oaktrees one has seen before. But that, of course, leads us back to the fuzzy sets we started from.

But still a more major problem with probability theory in the context of fuzzy semantics is its lack of a truth-functionality, which is, in our point of view, absolutely necessary to allow for a compositional treatment of language. The multiplication sometimes misunderstood as a probabilistic "conjunction" requires an independence assumption about its "conjuncts" which is in the context of fuzzy semantics less than easily justified. Reference is sometimes being made to a *probability logic* based on the Kolmogorov axioms in which independence assumptions are not made. Such a probability logic is in our point of view less than useful to our application, as it does not offer a closed conjunction operator.

The success of probability theory is, of course, undisputed here in its applications to syntactic language models. The scope of such models is a (stochastic) process by which syntactic elements end up in corpora. These are events in closed event-spaces that can easily be counted. We have outlined before that it is hard to define what sort of events one would exactly be looking for when modelling semantics probabilistically. But even if there was a possible definition, how could we justify any independence assumptions about semantic elements, when the fact that expressions have meaning exactly because they enter into functional relationships on the semantic level is the very basis of our endeavor to model semantics?

Using RMRS notation, the adjective phrase 'cold and rainy town' would be underspecified as $l_1 : \widetilde{\text{cold}}'(x_1), l_2 : \widetilde{\text{rainy}}'(x_2), l_3 : \widetilde{\text{town}}'(x_3)$ until a semantic model is invoked which determines that $x_1 = x_2 = x_3$ and that $l_1 = l_2 = l_3$, so we have $\widetilde{\text{cold}}'(x) \mathbin{\tilde{\wedge}} \widetilde{\text{rainy}}'(x) \mathbin{\tilde{\wedge}} \widetilde{\text{town}}'(x)$. It can be seen that the independence assumption is justified for the former, not for the latter, because if x is cold it may be (meteorologically) more likely to be rainy than if it was warm and if x is rainy it may be more likely to be cold than if it was dry. The effect will be that, syntactically, $l_1 : \widetilde{\text{cold}}'(x_1), l_2 : \widetilde{\text{rainy}}'(x_2)$ will be observed more often than $l_1 : \widetilde{\text{cold}}'(x_1), l_2 : \widetilde{\text{dry}}'(x_2)$ so, by conditioning syntactic probabilities appropriately, the language model may perform better.

This is due to the fact that the occurrences in the corpus will often mean $\widetilde{\text{cold}}'(x_1) \mathbin{\tilde{\wedge}} \widetilde{\text{dry}}'(x_1)$ so, by conditioning on more lexical information, the syntactic model is actually drawing in more semantic information, which is the reason why it performs better. In order to improve the language model, one would have to draw in more and more information to condition on, until one is effectively left with a model that includes a perfect treatment of semantics by conditioning every observable event on every other event making any independence assumption impossible. The model we are left with would be a probabilistic grammar in which we are multiplying numbers between zero and one, but where no independence assumption would be made. And that, of course, is exactly what the product t-norm does for us.

2.5 Fuzzy generalized quantification

2.6 Fuzzy generalized quantifiers

We have now defined fuzzy sets and relations, and their conjunctions and showed how they interact in fuzzy semantics to define vague natural language concepts like $\text{bald}'(x)$ or $\text{bald-millionaire}'(x)$. Now what about the meaning of a full proposition like 'Most millionaires are bald'? To define this, we need some way of accounting for the second level concept[1] behind 'most'. We shall see that 'most' is fundamentally different from 'millionaire', in that it denotes a *generalized quantifier*.

More particularly we will take the quantifiers denoting natural language concepts to be Lindström quantifiers, of type $\langle 1, 1 \rangle$ that fulfill the principles of conservativity, and extension. Without loss of generality we will also assume isomorphy. Thus the quantifiers we shall be interested in can be defined for the crisp case as follows:

Definition 8. A quantifier Q_X on domain X, is a mapping $Q : \mathcal{P}(X) \times \mathcal{P}(X) \mapsto \{\mathbf{F}, \mathbf{T}\}$ such that all of the following properties hold. For all possible restrictions and bodys

[1]after Frege's "Begriff zweiter Stufe"

$R, B \in \mathcal{P}(X)$

$$Q_X(R, B) = Q(R, R \cap B). \qquad \text{(conservativity)} \tag{2.24}$$

If $X' \in \mathcal{P}(X)$ is a subdomain of X, and we take $R, B \in \mathcal{P}(X')$, then

$$Q_{X'}(R, B) = Q_X(R, B). \qquad \text{(extensibility)} \tag{2.25}$$

Let $R, B \in \mathcal{P}(X)$ and let Y be a different domain than X. Further assume we have a bijection f from X to Y. Now

$$Q_X(R, B) = Q_Y(f(R), f(B)) \qquad \text{(isomorphy)} \tag{2.26}$$

So we will generally use a quantifier Q as a mapping, that has, as its first argument, a restrictor R and, as its second argument, a body B. This mapping ranges over truth values. It has been observed that quantifiers that represent denotations of natural language determiners always fulfill requirements **??** and **??**.

Requirement **??** amounts to the observation that the truth evaluation of a natural language quantifier only depends on the restriction and the intersection of the restriction with the body, never on elements of the body that do not fulfill the restriction. Requirement **??** states that, by fixing a domain X', choosing restriction and body as subsets of this domain, and applying a quantifier to them gives us a truth value that will stay the same, if we add elements to the universe X' to obtain an extended universe X. This directly follows from **??**, since the newly added element cannot have been part of the old restriction.

We assume requirement **??** without loss of generality, since we can always have an artificial domain in which individual elements have no inheren meaning. All elements must then be identified solely by predicates on them, i.e. instead of assigning an element k the inherent meaning that it refers to kitty, we will always denote kitty as the x which fulfills $\text{kitty}'(x)$.

Again, we can generalize to the fuzzy case:

Definition 9. A fuzzy quantifier Q_X on domain X, is a mapping $Q : \mathcal{P}(X) \times \mathcal{P}(X) \mapsto \{\mathbf{F}, \mathbf{T}\}$ such that all of the following properties hold. For all possible restrictions and bodys $R, B \in \mathcal{P}(X)$

$$Q_X(R, B) = Q(R, R \cap B). \qquad \text{(conservativity)} \tag{2.27}$$

If $X' \in \mathcal{P}(X)$ is a subdomain of X, and we take $R, B \in \mathcal{P}(X')$, then

$$Q_{X'}(R, B) = Q_X(R, B). \qquad \text{(extensibility)} \tag{2.28}$$

Let $R, B \in \mathcal{P}(X)$ and let Y be a different domain than X. Further assume we have a bijection f from X to Y. Now

$$Q_X(R, B) = Q_Y(f(R), f(B)) \qquad \text{(isomorphy)} \tag{2.29}$$

2.7 Conclusions

In this chapter, we have introduced the reader to fuzzy logic and linguistic vagueness. We have introduced the notion of ordering based semantics, and have shown how fuzzy logic might be used as a model of the ordering based semantics of natural languages. We have used different linguistic intuitions to identify a class of such models that are plausible and we have made a number of design choices that allow for a practical implementation and an empirical evaluation of such a model. Although we make absolutely no claim to the extent that our present model is necessary with respect to linguistic phenomena or optimal in any way, we do believe that, together with the firm justifications we offered and the extensive self-criticism made, this chapter represents an important first step towards such an optimal model.

Chapter 3

Experiment

In the previous section, we started out from the realization that strict bivalent decision boundaries do not adequately model the logic behind the semantics of vague expressions. We have proposed a model based on fuzzy logic which we claimed provides a more adequate treatment of such vague expressions.

In order to put this to the test, we have set up an experiment in which we ask subjects to place strict binary decision boundaries. Similar experiments have recently been taken up by Kees van Deemter et al. (van Deemter 2006*a*, van Deemter et al. 2006*a*,*b*, van Deemter 2006*b*) in the context of his work on natural language generation involving gradable expressions, and by Moxey & Sanford (2000, 1997) who concentrate on quantifiers.

In our experiment, we asked subjects about their agreement with statements like 'If a city has a population of 385412, it can be said to be a big city'. If a subject answers this question with 'no', but a question about a city with a 412536 population (which is the next larger data-point in a sorted sample) with 'yes', a boundary is hypothesized between those values that decides about the bivalent truth evaluation of a predicate big′.

As a null hypothesis we will use a model in which there is a representation for big′ in terms of a crisp set. As an alternative hypothesis we will use a model in which the representation is a fuzzy set $\widetilde{\text{big}}'$. As judges are asked to place strict bivalent boundaries, they choose a cutoff-point α at random and obtain a decision boundary as the preimage of α under the characteristic function as shown in Figure 3.1. In section 3.1 we will describe

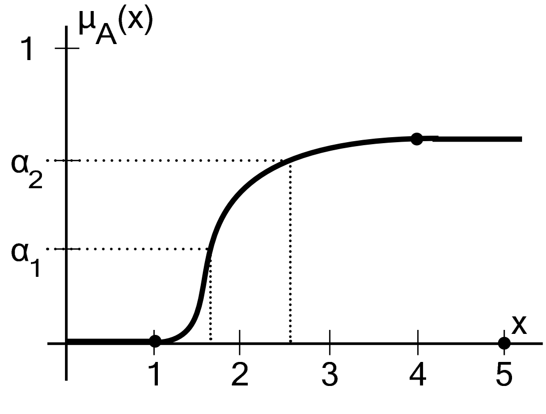


Figure 3.1: relationship between decision boundaries and orderings imposed by fuzzy sets

this experimental setup in detail, and we will analyze the results in section 3.2.

Although we will be looking at our specific experiment in isolation in this chapter, it should be noted that the software for this experiment was developed using our linguistic data modelling toolset which generalizes over our specific experiment and allows rapid-prototyping of software supporting similar experiments (possibly in different domains, with different vocabulary, etc). We believe that our major contribution does not lie within the specific findings resulting from this particular instantiation of our experiment, but rather in the provision of the software tools and statistic methods necessary to support future experiments of the general kind described herein.

We will also see that, as a side product of our statistic analyses, we will also get parameter values determining actual fuzzy sets characterizing the denotations of a number of interesting vague adjectives, which we will then put to use in the next chapter to produce an NLID.

3.1 Experimental setup

3.1.1 Background data and test-group

We asked people to judge the truth of different vague predicates about cities and skyscrapers. To this end, we collected data about cities in California from the U.S. census bureau (USCB 2000), providing the names of cities together with their global positions and their populations. Depending on the population of a city it may be described as *tiny*, *small*, *big*, or *huge*. We then used the global positions to find the geographic distance between two cities. Depending on this distance one city may be said to be *near* another city. Furthermore we used the global positions to link the census data with 10 year means of meteorological data from the U.N. IPCC (intergovernmental panel on climate change) (IPCC 2005). This provided 10-year average temperatures in degrees celsius which determine whether a city can be called *hot* or *cold* and the number of wet days in a year, which determines whether a city can be called *rainy* or *dry*. From this data, a random sample of 200 records was then taken to make the data easier to process.

Furthermore we used data about 26 skyscrapers from a 'TOP-TRUMPS' card-game, listing impressive skyscrapers around the world with their completion dates (*old* vs. *new*) and the number of floors (*big* vs. *small*).

Responses to our interactive questionnaire were collected in the period from JUN-26 to JUL-7 from volunteers responding to a call for participation advertised among personal friends of the author, as well as the RMRS mailing list and the local mailing list of the Cambridge NLIP group.

3.1.2 Design of the interactive questionnaire

These decision boundaries were obtained by asking sequences of questions generated from templates. For example the template for the temperature attribute of the city entity was *If a city had a year-round average temperature of #temp degrees celsius, it would be natural to call it a #ap city.* The templates used for each of the other attributes are given in the linguistic data model in appendix C. Questions were generated from this template using three different methods.

Referring expression generation Here the *#ap* placeholder would be resolved to a referring expression licensed by a context-free grammar. For the temperature example this grammar was defined as follows:

$$
\begin{aligned}
ap &\rightarrow adv\ adj, \\
adv &\rightarrow very, \\
adv &\rightarrow quite, \\
adv &\rightarrow rather, \\
adj &\rightarrow hot, \\
adj &\rightarrow cold.
\end{aligned}
$$

Appendix C gives the other grammars used here. Using these grammars, the questions are generated by randomly selecting a data-point to fill in for example for *#temp* and randomly expanding the grammar to resolve *#ap*. The final questions would look like this:

- *If a city had a year-round average temperature of 12 degrees celsius, it would be natural to call it a very cold city (yes/no)*
- *If a city had a year-round average temperature of 9 degrees celsius, it would be natural to call it a rather warm city (yes/no)*
- *If a city had a year-round average temperature of 36 degrees celsius, it would be natural to call it a rather cold city (yes/no)*
- ...

Open question Here *#ap* would be resolved to a gap to be filled in by the user. Seeing the denotation experiment as part of an iterative process for specifying the right query language, this feeds back relevant information about the lexical items that users would naturally want to use in the given linguistic context. This can be used to refine the specification of the experiment for another iteration.

- *If a city had a year-round average temperature of 12 degrees celsius, it would be natural to call it a _____ city (please fill in the gap)*
- *If a city had a year-round average temperature of 9 degrees celsius, it would be natural to call it a _____ city (please fill in the gap)*
- *If a city had a year-round average temperature of 36 degrees celsius, it would be natural to call it a _____ city (please fill in the gap)*
- ...

Lexical entry substitution Here the placeholder *#ap* is resolved to an adjective directly. These questions are always generated systematically using a binary search for a decision boundary over the data-sample sorted by the relevant attribute. The algorithm is somewhat reminiscent of the game, where someone thinks of a number (the decision boundary) in a certain range (as specified by the sample), and someone has to find the number by asking as few questions as possible. There is, of course, one such sequence of questions for each adjective to be trained. For *hot*, for example, the computer would first substitute the smallest value in the sample: *If a city had a year-round average temperature of 9 degrees celsius, it would be natural to call it a hot city (yes/no).* Since this will naturally be answered 'no', the computer will conclude that $\mu_{\widetilde{\text{hot}}'}(x)$ is nondecreasing in x.temp. For the parametric form $\mu_{\widetilde{\text{hot}}'}(x) = \mathbf{stdep}(x; \text{pop}, d, l, u)$, this already determines $d = +1$. Subsequently, the computer proceeds with a binary search: Initially the range in which the decision boundary lies is the full sample from the lowest, to the highest-ranking record. The next question about *hot* asks for the median of that sample. Given that we have a non-decreasing predicate, if the answer to that question is "yes", the range that remains to be searched is from the lowest-ranking element to the median. If the answer is "no", the new search range is from the median to the highest-ranking element. For non-increasing predicates the relation is inverse. This proceeds recursively, until two neighboring data-points receive different judgements, which leads the computer to record a decision boundary at that point. Assume a simplified example in which we are searching over the range of integers between zero and eight, and there was exactly one occurence of every such integer in the sample:

- *0 is a large number between 0 and 8: no*
- *5 is a large number between 0 and 8: yes*
- *2 is a large number between 0 and 8: no*
- *3 is a large number between 0 and 8: no*
- *4 is a large number between 0 and 8: no*

This would lead to a decision boundary at 5. This algorithm ensures that judgements are sampled with a higher resolution in the more "interesting" areas immediately surrounding the decision boundary, and that judgements are available that allow to hypothesize a specific decision boundary for each judge.

The user remains unaware of any systematic processes used, because the lexical items being trained and the methods by which questions are asked alternate randomly. This makes it harder for the user to memorize and artificially structure the answers, which would be counter-productive for an experiment aimed at the user's ad-hoc linguistic intuition.

The web-application used for administering this experiment was developed using our linguistic data modelling toolset, and the linguistic data model is given in appendix C. Subjects had access to the instructions given in appendix C.

3.1.3 Discussion and future work

At this point one may ask: Why is, for example, population taken to be the measure for smallness, as opposed to area, population density, or the existence of urban infrastructure such as a cathedral or underground transportation? In response to this we should mention that our approach to this experiment was that of linguistic data modelling. Here we assumed a database about a certain domain is already given, and an NLID needs to be developed based on that database. Under this view the experiment is merely collecting linguistic judgements supporting the construction of such an NLID, which is very different from the point of view that many linguists will naturally employ: Our aim is not the construction of a "Cambridge English Dictionary: Vague Adjective ⇒ Denotation". For such a project, the adjective *big* would have to be treated as polysemous, and an attempt should be made to make a comprehensive listing of senses such as *big_1* in which it is applied to cities, a sense *big_2* in which it is applied to buildings, etc and it would certainly need to be further subdivided into a sense *big_1_1* where it refers to a city's population, a sense *big_1_1* where it refers to a city's area, etc. From the point of view of linguistic data modelling however, we do not aim for any comprehensive treatment of any given lexeme beyond the information that is readily available in the database.

This also explains why we haven't given a more thorough justification of the choice of our background data. Judgements obtained about wet and rainy cities are certainly less useful for a broad-coverage dictionary when the sample of cities is restricted to California. From the point of view of linguistic data modelling again, this sample is perfectly acceptable if the aim is to construct an NLID specifically for a database of cities in California. As our experimental samples are drawn by a random number generator from all data that will be used in the NLID the samples are of a very high quality, in the sense that they are representative of the data used in the application.

The phrasings of the example questions also reflect this approach. Subjects were sometimes complaining that they were missing a frame of reference for their judgements. What does it mean for a city to be *near* another? Is this to be judged by European or by U.S. standards? Cambridge and Ely are close together for someone in Edinburgh planning a roadtrip, but not for a Cambridge student looking for accommodation. Such a frame of reference was not provided in the experiment. Rather users were asked to make an assumption and stick with it. Again, the overall goal was to collect data that is representative of what happens in the actual application. A user querying for a 'city near Cambridge' on an NLID will want the computer to make reasonable assumptions to answer this in just the same way as the experiment asks judges to make reasonable assumptions.

From this it gets obvious that, by taking into account vagueness, we often have to introduce new levels of ambiguity into our lexicon. But ambiguity is not the issue in our experiment. This is why we circumvented the issue altogether by employing our application-oriented point of view, obtaining judgements that are representative for a specific application, rather than judgements that are universally applicable.

As far as the composition of the test-group is concerned one major weakness is that most subjects reading the RMRS-mailing list or being members of the Cambridge NLIP group will be specialists and may therefore not be quite representative for either the userbase one might expect for a real-life NLID or the population of all speakers of English.

The major weakness of the experimental interface itself may be its inability to model the complex interactions of lexical items in referring expressions. The template-substitution approach makes it very hard to rule out semantically self-contradictory referring expressions like 'skyscraper with 8 floors'. A much more desirable solution would be to determine the denotation of the noun *skyscraper* first and, after having learned all nouns, using them correctly in referring expressions that experiment with adjectives modifying them.

item	N		κ_1	κ_2	MSE$_6$	MSE$'_6$
tiny city	26	3	0.56	-0.19	0.478	0.657
small city	25	5	0.32	-0.20	0.324	0.305
big city	26	0	1.00	-0.26	0.149	0.158
huge city	26	0	1.00	-0.91	0.060	0.117
hot city	18	0	1.00	-0.49	0.209	0.530
cold city	18	2	0.58	-0.28	0.288	0.400
dry city	13	2	0.43	-0.53	0.211	0.216
rainy city	13	1	0.69	-0.45	0.257	0.243
near city	23	0	1.00	-0.21	0.068	0.680
small skyscraper	14	1	0.71	-0.36	0.250	0.473
big skyscraper	14	0	1.00	-0.41	0.114	0.236
old skyscraper	13	2	0.43	-0.53	0.162	0.424
new skyscraper	13	0	1.00	-0.44	0.282	0.569

Figure 3.2: Some statistics about the various vague concepts

Furthermore the experiment is quite sensitive to the subject answering the first question (which decides about whether the function will be taken as nonincreasing or nondecreasing) correctly. A more direct way of establishing this parameter-value might turn out to perform better.

3.2 Results and statistics

3.2.1 Plausibility and crisp sets

The number N of judgements received for each word are given in Figure 3.2. Figures 3.3 and 3.4 show the decision boundaries as placed by the judges in scatter plots.

Null Hypothesis 1. There is no agreement among subjects about whether a vague predicate should be considered non-increasing or non-decreasing in a measurement the predicate's truth is a function of which.

Alternative Hypothesis: As stated in Hypothesis 1, a fuzzy set that is the denotation of a vague predicate can always clearly be judged as nonincreasing or nondecreasing.

The second column of Figure 3.2 shows how many judges disagreed with the majority opinion for each of the words, and the same agreement κ_1 measured in terms of Cohen's Kappa.

In our opinion, the null hypothesis can be rejected for *big*, *huge*, *hot*, and *rainy* cities as well as for cities *near* other cities, and for *small*, *big*, and *new* skyscrapers. The failure of the experiment to contradict this null hypothesis for the other words may be attributed to the fact that, when asked about a skyscraper with 9 floors or a skyscraper built in 1843, people might judge that this cannot be called a *small* or *old* skyscraper, not because of the vague adjective in question, but because of the noun: Such buildings would not normally be called skyscrapers. The difficulty here is that, in the absence of a model to deal with the complex grammatical interactions, a way was needed to isolate

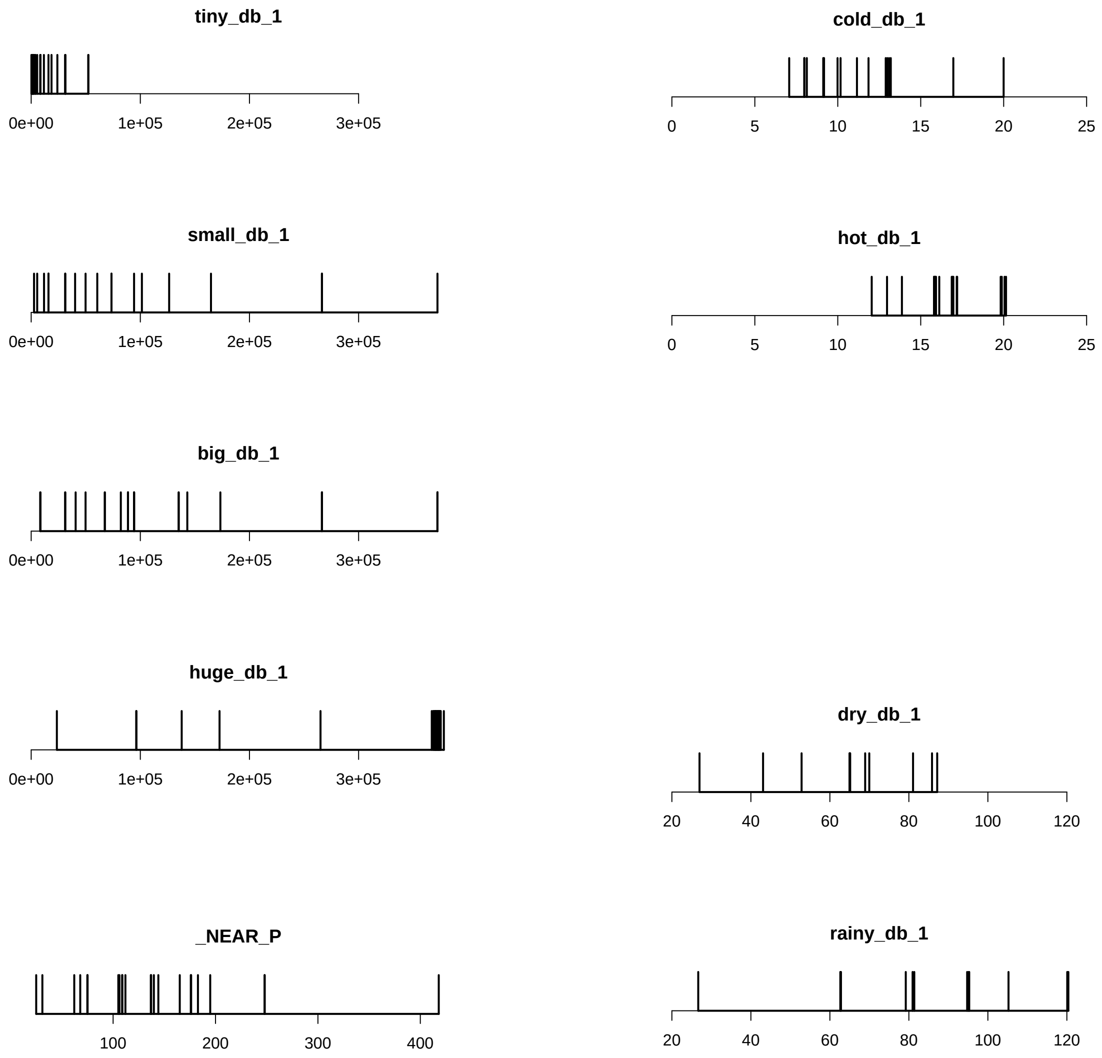


Figure 3.3: Scatter plots of the boundary placements in the cities domain

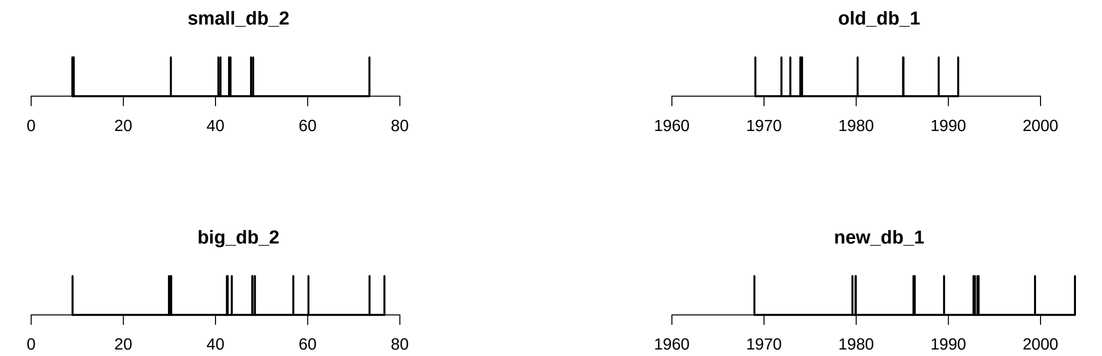


Figure 3.4: Scatter plots of the boundary placements in the skyscrapers domain

the contribution of the vague adjective. This was done by introducing, in the question's if-part a hypothetical scenario under which a building with 9 floors is called a skyscraper. Although the instructions clearly asked the subjects to focus on the adjectives, and to proceed under the hypothesis presented in the if-parts of the questions as far as the noun is concerned, this may explain why the experiment failed for *tiny* and *small* city, and *old* skyscraper, but it cannot possibly account for *dry* city and *cold* city.

Since the algorithm by which a subject is interrogated for a decision boundary depends on this judgement to be made correctly, the data in question had to be removed from the results for subsequent tests.

3.2.2 Testing the model of fuzzy semantics

Null Hypothesis 2. No subjects place the decision boundary for *tiny* or *big* cities higher than their own or some other subject's decision boundary for *small* or *huge* cities respectively.

Alternative Hypothesis: As stated in hypothesis 4, decision boundaries as well as fuzzy sets for different speakers may be contradictory, but each speaker is self-consistent about them.

Out of 460 pairs of *tiny/small* judgements across judges, we found that the decision boundary for *tiny* was higher than that for *small* in 67 cases (14.56%). Out of 676 pairs of *big/huge* judgements, we found that the decision boundary for *big* was higher than that for *huge* in 65 cases (9.61%). We can clearly reject the null hypothesis.

Null Hypothesis 3. Some subjects place the decision boundary for *tiny* or *big* cities higher than their own decision boundary for *small* or *huge* cities respectively.

Alternative Hypothesis: As stated in hypothesis 4, decision boundaries as well as fuzzy sets for different speakers may be contradictory, but each speaker is self-consistent about them.

Out of 20 judges who made valid judgements about both *tiny* and *small*, none placed a decision boundary for *tiny* higher than for *small*. Out of 26 judges who made valid judgements about both *big* and *huge*, none placed a decision boundary for *big* higher than for *huge*. The null hypothesis can clearly be rejected.

Null Hypothesis 4. There is a high level of agreement $\kappa \geq \kappa_\theta$ among subjects about the bivalent truth valuation of a vague predicate across all measurement points.

Alternative Hypothesis: Vague predicates cannot be modeled adequately by bivalent decision boundaries.

Figures 3.5 and 3.6 show the agreement of subjects about the bivalent valuation of the predicates as a function of the underlying measurement. This function is defined as follows:

$$\kappa(x) = \frac{P_A(x) * P_E}{1 - P_E}$$

Since the judgement is about a bivalent valuation P_E is 0.5. Furthermore

$$P_A(x) = \frac{\#_<(x) * (\#_<(x) - 1) + \#_>(x) * (\#_>(x) - 1)}{N(N-1)},$$

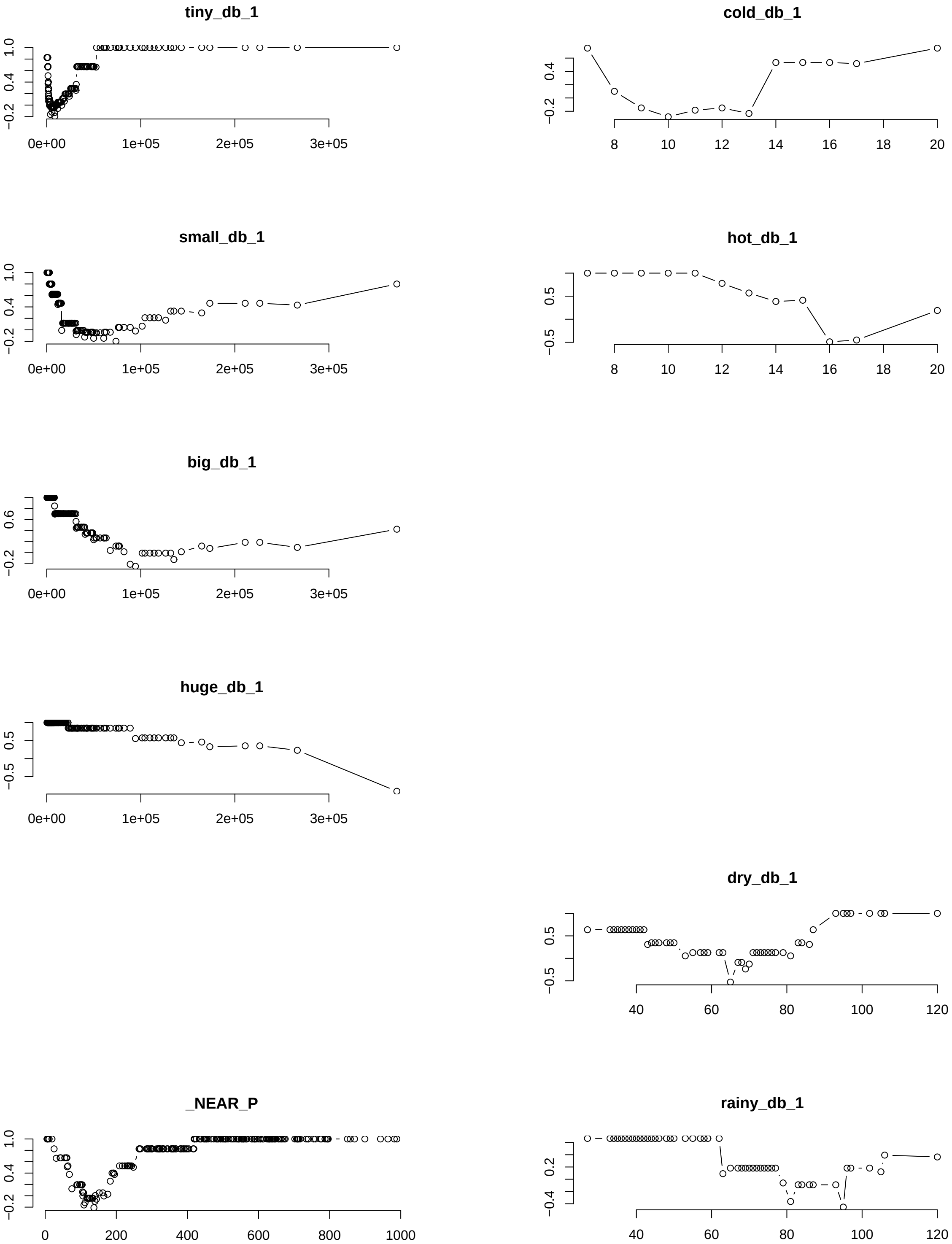


Figure 3.5: $\kappa(x)$ as functions of the measurements in the cities domain

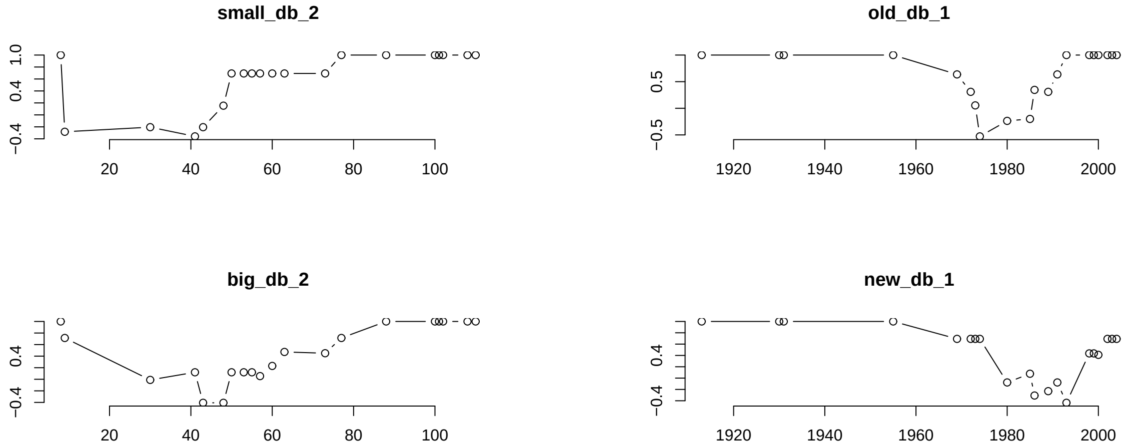


Figure 3.6: $\kappa(x)$ as functions of the measurements in the skyscrapers domain

where

$$\begin{aligned} \#_<(x) &= |\{j|x < b_j\}|, \\ \#_>(x) &= |\{j|x > b_j\}|, \end{aligned}$$

and j is a judge placing a decision boundary at b_j. Figure 3.2 shows the minimal level of agreement $\kappa_4 = \min_x \kappa(x)$ obtained in the experiment for any x.

The experiment contradicts the null hypothesis for all words and for any choice of $\kappa_\theta > -0.19$, i.e. with overwhelming confidence.

Null Hypothesis 5. Suppose the denotation of a vague predicate was a fuzzy set of the form $\mathbf{stdep}(x; att, d, l, u)$ as claimed earlier. There would have to be a high level of agreement $\kappa \geq \kappa_\theta$ among subjects about the bivalent truth valuation of a vague predicate across a set of measurements with $x.att \leq l$ or $x.att \geq u$. Only a minor proportion $p \leq p_\theta$ of typical values for x in a random sample falls within that range.

Alternative Hypothesis: As stated in hypothesis 2, a region can always be clearly identified in which the set does not behave like a crisp set.

Figure 3.7 shows the relevant values: Here $p_5(\kappa_5 \geq \kappa_\theta)$ is the proportion of the sample that falls within a threshold $\kappa_5 \geq \kappa_\theta$, and $\kappa_5(p_5 \geq p_\theta)$ is the minimal level of agreement obtained for a portion of the sample that falls within a threshold $p_5 \geq p_\theta$.

In our opinion, the data justifies rejecting the null hypothesis for *small*, *big*, *huge*, *hot*, and *dry*, cities, cities *near* others, *small*, *big*, and *old* skyscrapers.

The experiment seems to fail on *rainy* and *cold* cities. This might be attributed to the fact that the sample of cities used in the experiment was taken from a database of all cities in California and is missing good prototypes for rainy and cold cities.

Null Hypothesis 6. Suppose the denotation of a vague predicate $\widetilde{A}$ was a fuzzy set. Then, by forcing a strict binary decision boundary, a judge would obtain a cutoff point α at random, and place the decision boundary as the preimage of α under the characteristic

item	$p_5(\kappa_5 \geq 0.50)$	$p_5(\kappa_5 \geq 0.75)$	$p_5(\kappa_5 \geq 1.00)$
tiny city	0.31	0.13	0.13
small city	0.56	0.38	0.18
big city	0.75	0.50	0.49
huge city	0.97	0.93	0.67
hot city	0.58	0.50	0.42
cold city	0.50	0.16	0.08
dry city	0.39	0.18	0.18
rainy city	0.47	0.02	0.02
near city	0.83	0.74	0.53
small skyscraper	0.75	0.40	0.40
big skyscraper	0.45	0.35	0.35
old skyscraper	0.68	0.59	0.59
new skyscraper	0.59	0.27	0.27

item	$\kappa_5(p_5 \geq 0.25)$	$\kappa_5(p_5 \geq 0.50)$	$\kappa_5(p_5 \geq 0.75)$
tiny city	0.65	0.05	< -0.1
small city	0.80	0.60	-0.05
big city	1.00	0.65	0.40
huge city	1.00	1.00	0.80
hot city	1.00	0.75	0.35
cold city	0.50	0.50	< -0.1
dry city	0.60	0.30	0.05
rainy city	0.65	0.35	0.05
near city	1.0	1.0	0.65
small skyscraper	1.00	0.65	0.65
big skyscraper	1.00	0.45	0.05
old skyscraper	1.00	1.00	0.30
new skyscraper	1.00	0.65	-0.10

Figure 3.7: Statistics for null hypothesis 5

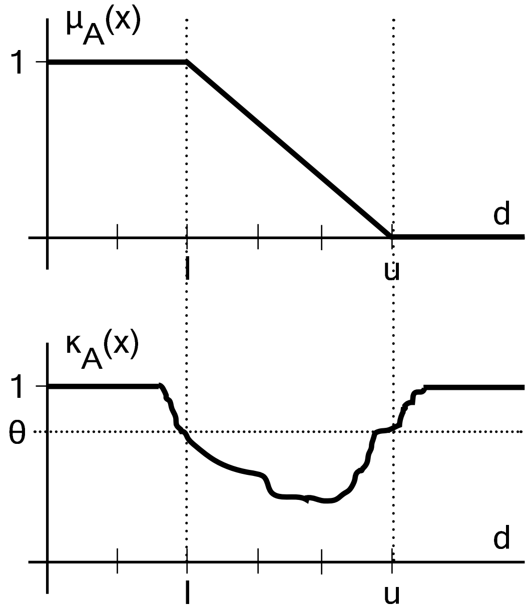


Figure 3.8: relationship between observed agreement and hypothesized fuzzy sets

function. The agreement between judges simulated in such a way cannot outperform a baseline in which judges are simulated to choose their decision boundaries at random from the range of the sample.

Alternative Hypothesis: As stated in hypothesis 3, our assumed parametric form of fuzzy sets representing vague predicates yields an adequate model for human intuition about vagueness.

The mean squared error between the observed function $\kappa(x)$ and a function $\kappa'(x)$ obtained from such a simulation

$$\mathrm{MSE}_6 = \mathcal{E}_x\{(\kappa(x) - \kappa'(x)')^2\}$$

is shown in Figure 3.2, together with the baseline MSE_6'. The fuzzy sets in the simulation were defined by setting the fuzzy region to the region in which κ is smaller than a threshold $\theta = 0.75$. Figure 3.8 depicts this relationship between observed agreement and hypothesized fuzzy set.

In our opinion, the null hypothesis can be rejected for *tiny*, *huge*, *hot* and *cold* cities, for cities *near* others and for *small*, *old* and *new* skyscrapers. The null hypothesis cannot be rejected for *small* and *rainy* cities. This could be attributed to the fact that the baseline happens to perform relatively well, given that most cities in the sample are small. For *rainy* cities, the sample may again be lacking the right prototypes.

3.2.3 Discussion and future work

In the previous section we have already identified some weaknesses in our particular experiment which were possibly inevitable given the limited time and personal resources of this project. We had to invest a considerable amount of work into the development of the software infrastructure necessary to support this experiment, which is why we view this as the major deliverable rather than the particular data we collected in our preliminary experiment. With this infrastructure in place it should be possible to repeat the experiment using more carefully compiled sample data and a better test group.

Nevertheless the results obtained from our preliminary experiment are quite encouraging in that they serve as a proof-of-concept for our theoretic models, software infrastructure and experimental design. In this section specifically we have demonstrated how data obtained from the experiment can be statistically analyzed to provide evidence about the adequacy of models about fuzzy semantics such as our own and the limits of software based on such models such as the NLID we will describe in the next chapter.

The present work could also be extended to compare alternative parametric forms of fuzzy sets, such as Zadeh's "S-shaped" function and alternative definitions of target attributes (direct measurements, percentiles, normalized ranks, ...).

3.3 Conclusions

In this chapter we described the overall design of our experiment and the statistical analyses required to draw conclusions from the data about the adequacy of models of fuzzy semantics.

Based on a small-scale preliminary instantiation of this experiment we have reason to be optimistic about our general framework. The data clearly supports our claim that strict bivalent decision boundaries are inadequate as models of vague expressions, and that the model of fuzzy semantics we suggested is a promising alternative.

As a side product of our statistic analyses, we have also extracted good parameter values determining actual fuzzy sets characterizing the denotations of a number of interesting vague adjectives, which we can put to use in the next chapter to produce an NLID.

Chapter 4

A fuzzy natural language database interface

In the previous chapters we have introducted our model of fuzzy semantics and have then moved on to provide some empirical evidence in support of this model. In this chapter we will show how we could put those theoretic insights to use to produce a working natural language interface to a database (NLID) that uses the fuzzy sets derived from the experiments in the previous section to produce orderings of records in a database ranked according to the degree to which they fulfill our intuitions behind expressions involving vague adjectives like ‘small city’ or ‘rainy city near San Francisco’.

For general background on these applications the reader is referred to the introductions by Androutsopoulos (1995), Androutsopoulos (2000), and Copestake & Spärck-Jones (1990).

Figure 4.1 shows a screenshot of our demo database interface. Just as the software running our interactive questionnaire this database interface was developed using our linguistic data modelling toolset, so although our specific instantiation of this database interface operates in the cities/skyscrapers domain introduced in the previous chapter, we can adapt this to another domain as a matter of minutes.

4.1 The query language

4.1.1 Some example queries in English

A query like ‘big city’ is resolved in our NLID to a result set in which all cities that are undoubtedly *big* appear first, ordered randomly. In the region where it is unclear whether or not a city may be big, the records will be in order of decreasing population. The last group of records will be those cities which are definitely not *big*, again in random order.

A query like ‘big rainy city’ will retrieve a result set with a slightly more complex structure, since we have two criteria. Essentially the query ‘rainy city’ is analogous to the above example. The result set for ‘big rainy city’ is now a combination of the two in which the ordering in ‘rainy city’ decides the ordering of records that are ties in ‘big city’ and vice versa. So for all the cities that are undoubtedly *big* the records will now appear in an order that is not random, but dictated by the records’ fulfillment of the *rainy* criterion.

The query ‘city near San Francisco’ basically involves two entities. In this case all fields

Natural Language Database Demo Interface

Query: rainy city [Submit]

dof	mainid	x4.placeid	x4.placename	x4.lat	x4.long	x4.pop	x4.temp	x4.wet
0.240	380	380	Klamath	846091	-2164148	827	11	124
0.200	26	26	Arcata	857517	-2165620	15197	11	120
0.200	49	49	Bayview	859313	-2167279	1318	11	120
0.200	74	74	Blue Lake	857301	-2164061	1235	11	120
0.200	177	177	Crescent City	842049	-2167661	4380	10	120
0.200	178	178	Crescent City North	841882	-2167944	3853	10	120
0.200	185	185	Cutten	859270	-2166686	1516	11	120
0.200	256	256	Eureka	858783	-2166925	27025	11	120
0.200	465	465	McKinleyville	856026	-2166226	10749	11	120
0.200	525	525	Myrtletown	858894	-2166461	4413	11	120
0.200	791	791	Trinidad	854203	-2166688	362	11	120
0.200	836	836	Westhaven-Moonstone	854478	-2166003	1109	11	120
0.100	255	255	Etna	847195	-2144884	835	9	110
0.100	265	265	Ferndale	862550	-2168757	1331	12	110
0.100	278	278	Fort Jones	844616	-2143959	639	9	110
0.100	854	854	Willow Creek	857267	-2158367	1576	10	110
0.090	210	210	Dorris	838365	-2127904	892	8	109
0.070	828	828	Weaverville	859247	-2145862	3370	9	107
0.060	501	501	Montague	842514	-2138544	1415	9	106
0.060	520	520	Mount Shasta	849582	-2134794	3460	9	106
0.060	794	794	Tulelake	838558	-2120135	1010	8	106
0.060	829	829	Weed	847941	-2135905	3062	9	106
0.060	873	873	Yreka	842486	-2140322	6948	9	106
0.050	279	279	Fortuna	862440	-2166645	8788	11	105
0.050	345	345	Humboldt Hill	859990	-2167499	2865	11	105
0.050	349	349	Hydesville	863088	-2165728	1131	11	105
0.050	596	596	Pine Hills	859863	-2166845	2947	11	105
0.050	645	645	Rio Dell	863923	-2166064	3012	11	105
0.040	215	215	Dunsmuir	851135	-2134012	2129	9	104
0.040	277	277	Fort Bragg	882398	-2160768	6078	12	104
0.040	463	463	McCloud	850764	-2131661	1555	8	104
0.020	323	323	Hayfork	862648	-2148941	2605	10	102
0.020	426	426	Lewiston	860460	-2143299	1187	10	102
0.020	638	638	Redway	870610	-2161055	1212	11	102

(c) Copyright 2006 by Richard Bergmair.

Figure 4.1: database demo interface

retrieved as part of a join statement are also returned in the result set. The result set will have a group of fields associated with the city which is near San Francisco, and ordered in exactly the same way as if 'near San Francisco' was a simple one-place predicate like 'big', but in addition each record will have a group of fields pertaining to San Francisco. In a query like 'city near a rainy city' this allows the user to see not only data about the city near a rainy city they asked for but also data about the rainy city.

4.1.2 Some example queries in SQL

Our database interface assumes that nouns like *city* or *town* can uniquely be assigned to an entity in the database. To resolve the query 'city', for example, there may be an entity `place` so that this query can retrieve all records x where $x \in$ `place`, i.e.

$$\texttt{SELECT } x.\texttt{*} \texttt{ FROM place}$$

Proper nouns like *San Francisco* can be resolved directly to a value in a string attribute in the database. The noun *San Francisco*, in particular, may appear in a `placename` attribute associated with the `place` table. The query 'San Francisco' therefore would retrieve all $x \in$ `place` where x.`placename` = *San Francisco*, i.e.

$$\texttt{SELECT } x.\texttt{*} \texttt{ FROM place } x \texttt{ WHERE } x.\texttt{placename} = \textit{`San Francisco'}$$

Adjectives like *tiny*, *small*, *big*, and *huge*, on the other hand, are assigned to an attribute within a table, in our case the `pop` attribute of the `place` table. The query 'big city' would retrieve all records $x \in$ `place` where $\mu_{\widetilde{\mathrm{big}}'}(x) > 0$, ordered by degree of fulfillment, where $\mu_{\widetilde{\mathrm{big}}'}(x) = \mathbf{stdep}(x; \mathrm{pop}, d, l, u)$ for the d, l, and u determined in the experiment. Of course, different attributes can be combined conjunctively in one query, so that 'big dry city' retrieves all $x \in$ `place` ordered by $\mu_{\widetilde{\mathrm{big}}'}(x) \mathbin{\widetilde{\wedge}_*} \mu_{\widetilde{\mathrm{dry}}'}(x)$. Conceptually, the SQL statement looks like this:

$$\texttt{SELECT } x.\texttt{*},\ \mu_{\widetilde{\mathrm{big}}'}(x) \mathbin{\widetilde{\wedge}_*} \mu_{\widetilde{\mathrm{dry}}'}(x) \texttt{ AS } \mu \texttt{ FROM place } x \texttt{ WHERE } \mu > 0 \texttt{ ORDER BY } \mu \texttt{ DESC}$$

The word *near* is different from these adjectives in that it is relational. Again the denotation is of the form $\mu_{\widetilde{\mathrm{near}}'}(z) = \mathbf{stdep}(z; \mathrm{distance}, d, l, u)$, but this time z resolves to the pseudo-entity `refnear` $\subseteq$ `place` $\times$ `place`, so we effectively get $\mu_{\widetilde{\mathrm{near}}'}\big((z_x, z_y)\big) = \mathbf{stdep}\big((z_x, z_y); \mathrm{distance}, d, l, u\big)$. The query 'dry city near a rainy city' would retrieve all $(x, z, y) \in$ `place` $\times$ `refnear` $\times$ `place` where z is of the form (z_x, z_y) and a join is formulated so that z_x corresponds to x and z_y corresponds to y, ordered by $\mu_{\widetilde{\mathrm{dry}}'}(x) \mathbin{\widetilde{\wedge}_*} \mu_{\widetilde{\mathrm{near}}'}(z) \mathbin{\widetilde{\wedge}_*} \mu_{\widetilde{\mathrm{rainy}}'}(y)$.

Conceptually, the SQL statement looks like this:

$$
\begin{array}{rl}
\texttt{SELECT} & x.*, \\
 & z.*, \\
 & y.*, \\
 & \mu_{\widetilde{\text{dry}}'}(x)\ \widetilde{\wedge}_*\ \mu_{\widetilde{\text{near}}'}(z)\ \widetilde{\wedge}_*\ \mu_{\widetilde{\text{rainy}}'}(y)\ \texttt{AS}\ \mu \\
\texttt{FROM} & \texttt{place}\ x, \\
 & \texttt{refnear}\ z, \\
 & \texttt{place}\ y \\
\texttt{WHERE} & x.\texttt{placeid} = z.\texttt{placeid}\ \texttt{AND} \\
 & z.\texttt{fkplaceid} = y.\texttt{placeid}\ \texttt{AND} \\
 & \mu > 0 \\
\texttt{ORDER BY} & \mu\ \texttt{DESC}
\end{array}
$$

These basic types of reference that can be made in a natural language query can, of course, be combined as in 'big city near a big city near San Francisco' or 'big but also rainy city which is near a small dry town'.

4.1.3 Discussion & future work

From the descriptions given in this section it gets apparent that natural language queries may quickly develop into quite complex join-statements. By nature of its definition, a database application with a natural language interface will face a much wider variety of complex queries than the usual "canned" queries resulting from a graphical interface for instance. Therefore a database backend for use with an NLID can be quite demanding to administrate, as the necessary index-structures need to be in place to support join statements that cannot easily be forseen at development time.

In our preliminary experiments with our natural language database interface, we found out that even a query like 'city near San Francisco' in a database of cities in California takes a disappointingly long time to finish. This offers a point of departure for work on database optimization and performance evaluation for NLIDs.

At this point our database interface is a rather rudimentary proof-of-concept prototype. For a serious deployment it would have to be interfaced, for example, from stored procedures so that the system can be used by application frontends. (In our prototype the demo interface is its own frontend) Furthermore, it currently lacks support for adverbs which would have to be modelled by fuzzy quantifiers.

4.2 Prototype design

4.2.1 ERG-based semantic analysis

So far we have given very conservative examples, always resorting to syntactically very simple expressions like 'big rainy city near a small city', but, as a matter of fact, our database interface is quite robust to syntactic variations and would handle 'a city which is big and also rainy and near a city that is small' just as well. It ows this robustness to the broad coverage English Resource Grammar (ERG) which processes the input queries through the Linguistic Knowledge Builder (LKB) and returns a semantic representation of the query based on Minimal Recursion Semantics (MRS). For further details about MRS, the reader is referred to Copestake et al. (1999) and Copestake (2004). For details on the ERG and the LKB, the reader is referred to the DELPH-IN system homepage (DELPH-IN 2006). In this section we will concentrate on how exactly these components are configured and put to use in the specific context of our NLID.

Our system uses a special version of the ERG which is based on a future release currently under development and has an augmented interface to the LKB which enables calling into the LKB to read a phrase for the parser from a file. It uses a special output-method for the LKB which writes XML-based scoped-MRS structures into an output file[1].

Our own database-interface is entirely written in Python. The Python script interfaces the LISP-based LKB by calling it in a pipe and executing commands from its tty-interface, thereby effectively implementing a "LISP speaker" as an interface towards the "LISP listener" provided by the LKB. This component has an interface into the python-application which allows for user-friendly integration of LKB-functionality into Python programs. Furthermore the interface employs an XML-parser for the XML-MRS format, so our Python environment can effectively handle scoped MRS structures. This interface between Python and LKB/ERG/MRS was designed to be highly reusable to support future rapid-prototyping involving these DELPH-IN components.

The ERG-lexicon needs to be extended for each database domain to be used. Most importantly, the proper names have to be added, which is why our linguistic data modelling tool supports the automatic construction of a TDL file from a database. This way the value 'San Francisco' in the placename of our example database automatically becomes the following TDL type:

```
sanfrancisco_db_1 := n_proper_city_le &
  [ STEM < "San", "Francisco" >,
    SYNSEM [ LKEYS.KEYREL.CARG "sanfrancisco_db_1_rel",
             PHON.ONSET con ] ].
```

This automatically generated extension lexicon is loaded in our adapted ERG together with the standard lexicon.

Here it should be pointed out that the LKB has been designed to work with reasonably large lexica, so that there should be no problem for example with every city of the world becoming a lexical entry. The most elegant solution here would be to include those in the lexicon using the `lexdb` lexical database interface which loads a lexicon into the LKB directly from a PostgreSQL backend. However this wasn't necessary to get to the proof-of-concept level of implementation we aimed for in the current project.

4.2.2 MRS to SQL translation

After running the query expression through the ERG, the set of possible scoped MRSs is filtered through a disambiguation method working under the assumption that a query

[1]Thanks to Ann Copestake for making the necessary adaptations to these DELPH-IN tools

refers to exactly one table. In our example database there are two senses of *big*. The sense `big_db_1` applies to populations in the `place` table, and the sense `big_db_2` applies to the number of floors in the `skyscraper` table. Since the noun `city_db_1` can refer only to the `place` table, not to the `skyscraper` table, the word 'big' in 'big city' is automatically resolved to `big_db_1`, and all MRSs are filtered that contain an EP making reference to `big_db_2`.

If multiple scoped MRS structures are still valid (for example if *big* was ambiguous between a sense in which it applies to a city's population and a sense in which it applies to a city's area, or as a result of scope ambiguity) each will be translated into SQL separately, and the user will see all possible result sets.

At this point our demo interface employs a very simple mechanism for translating MRS into SQL: Each EP is considered in turn. Whenever an EP refers to a lexical item that has an interpretation with respect to the database, its denotation in the form of a fuzzy set is added to a fuzzy conjunction whose degree of fulfillment determines the ranking of the result set. The MRS-representation is thus resolved to a flat semantic representation, which suffices for our simple proof-of-concept prototype.

4.2.3 Discussion and future work

In this section we have described a design that allows for adequate treatment of a wide range of natural language queries through the use of the ERG as a linguistic backend.

Instead of using the LKB for running the ERG on the query phrases, a promising approach that might be explored in future work would be using the PET which is specifically designed as a runtime environment (as opposed to the LKB which is primarily a development environment).

Another obvious point of departure for future work would be to make full use of the MRS-based semantic representation produced by the ERG, rather than using a flat semantic representation. This would seem to require a proper treatment of fuzzy quantifiers which was outside the scope of our project.

4.3 Conclusions

In this chapter we were able to take our model of fuzzy semantics to a proof-of-concept level by putting it to use in an NLID that produces orderings of records in a database ranked according to the degree to which they fulfill our intuitions behind expressions involving vague adjectives like 'small city' or 'rainy city near San Francisco'.

We showed that, by relying on the ERG as a broad-coverage English grammar to provide the semantic analysis of our queries, we could gain an elegant design in which the database-interface itself can be implemented independently from any syntactic language model.

Chapter 5

Concluding remarks

In this work we have developed ordering based semantics as a radically different approach to semantics and have showed how fuzzy sets can be used as intermediate representations for the semantics of vague natural language concepts. We have used linguistic considerations to identify from the family of fuzzy logics that which best fits our modelling needs in fuzzy semantics.

We then went on to provide empirical evidence in support of our theory of fuzzy semantics. To this end we have developed a piece of software administering an interactive questionnaire over the web. By using human subjects we were able to collect data about intuitive judgements and draw conclusions about the adequacy of our model, based on some statistic methods we developed. From our preliminary instantiation of the experiment, we have reason to be optimistic about our theory of fuzzy semantics as well as our design of the experiment.

Next we moved on to put these theoretic insights to use in a natural language interface to a database (NLID). Again some extensive software development was involved in this stage. In this report we have given an overview of how exactly our system produces an ordering of records in a database, ranked according to the degree to which they fulfill our intuitions about queries involving vague adjectives like 'rainy city' or 'small town near San Francisco'.

Although the description of our work in this report concentrated on a closed example domain involving cities and skyscrapers, we believe that our insights generalize well to different applications. As far as the software development is concerned we actually took the problem to a meta-level by developing a toolkit that supports rapid prototyping of NLIDs in domains characterized in a linguistic data modelling (LDM) language of our own design.

At the end of the day we can only perhaps claim to have made a small contribution to an exciting field that has seen far too little research by now, so the vision of universal ordering based information access as enabled by the expressive power of our vague language may remain science fiction at this point, but hopefully we have convinced the reader in that it should perhaps be considered rather science than fiction.

Appendix A

Related work

In this section we will try to put the present work into some context of historic as well as current research in related fields. Most notably the present work follows up on a simpler framework we developed for the syntax-driven analysis of context-free languages with respect to fuzzy relational semantics (Bergmair 2006). This prior work was also presented at a major conference to the fuzzy systems community (Bergmair & Bodenhofer 2006). As far as natural language interfaces to databases are concerned, the reader is referred to the introductions by Androutsopoulos (1995), Androutsopoulos (2000), and Copestake & Spärck-Jones (1990). As far as our linguistic toolset, including the LKB and ERG, is concerned, the reader is referred to the DELPH-IN system homepage (DELPH-IN 2006). The following sections will focus on the theme of fuzzy semantics.

A.1 Fuzzy semantics

Lotfi Zadeh's landmark paper (Zadeh 1965) is probably the starting point of fuzzy logic as we know it today. Although the importance of everyday quantitative expressions for the mathematical analysis of vague concepts was recognized earlier, for example by (Sheppard 1954) in the domain of quantitative research methodology, Zadeh was the first to propose mathematical tools to cope with fuzziness on a broad scale. At the time of their inception, Zadeh's ideas were well received in the systems engineering community and developed into a rich toolset addressing the pressing needs of engineers to cope with imprecisely defined concepts in system specifications. (See Gaines & Kohout (1977) for an exposition of early work in the field).

In the course of the first rush of euphoria, formal language theory, being part of the body of engineering wisdom that was generically made subject to "fuzzification" during that time, was taken into the fuzzy domain by Lee & Zadeh (1969). However the idea of a fuzzified formal language theory did not receive much attention, when fuzzy systems were only just beginning to be successfully applied to simple control-tasks and formal language theory was of interest only to fields like compiler construction that naturally had little use for vague concepts.

One would assume that fuzzy logic, offering itself to linguists as a model of vague concepts and to cognitive psychologists as a model of degree-based reasoning, should have played a role in those fields in the years to come, but both remained widely unaware of the developments that took place in the engineering world up until the late 1970s, when Artificial Intelligence saw its rise as a major interdisciplinary field of study bringing together engineers with computer scientists, linguists and cognitive psychologists. The earliest mention of Zadeh's fuzzy logic in the linguistic literature is, to the best of the author's knowledge, made by Lakoff in the mid 1970s (see Parret 1974, p. 196), but the topic never received (possibly never deserved) much attention at a time when linguists were still struggling to work out the fundamentals of vagueness with bivalent decision boundaries, let alone think about what is, from a modelling perspective, an infinitude of them.

Furthermore, Artificiall intelligence drove a historically unparalleled interest in application-oriented meaning representation. Considerable work on fuzzy meaning representation schemes was carried out by Goguen (1974), who also attempted to build a fuzzy SHRDLU, a robot capable of carrying out commands input in natural language in the domain of a fuzzy microworld (Goguen 1975). Zadeh proposed a fuzzy meaning representation scheme for natural languages as well (Zadeh 1978). However, these representation schemes were mainly concerned with meaning as such, rather than meaning in relation to natural languages. Later, Zadeh presented test-score semantics (Zadeh 1981, 1982*b*) in an approach to bridge the gap between natural language representation and his fuzzy meaning representation. However his technique was never deployed in a wide-coverage language model.

Another important development of the mid 1970s was that, besides fuzzy logic, other methods of soft computing came along, such as neural networks and genetic algorithms. Fuzzy logic distinguished itself by putting renewed emphasis on the motivations that originally gave rise to its inception – the vaguely defined categories of reasoning employed by humans in natural language. It was realized that humans use natural language expressions to refer to those categories where fuzzy logic used sets defined in terms of numeric valued characteristic functions. This lead to the inception of the "linguistic variable" (Zadeh 1975*a*,*b*,*c*), a formal tool which makes explicit the correspondence between these two denotational variants of vague concepts. Engineers following the new paradigm were no longer free to pick fuzzy sets at will, but they had to bear in mind that fuzzy sets are meant to resemble meanings of natural language expressions. This aspect has only recently experienced renewed attention in an attempt to come to grips with what exactly it means for a linguistic variable to be interpretable in terms of natural language concepts (Bodenhofer & Bauer 2003, De Cock et al. 2000, De Cock & Kerre 2002).

After the early days of applying fuzzy logic to each and every conceivable problem that inspired the artificial intelligence community back in the 1970s, fuzzy logic experienced some decline in popularity in the decades that followed and matured to become the subject of major work on the foundations of mathematics, mainly carried out in eastern Europe, most notably the discovery of fuzzy logic as a generalization of classic logic that preserves its property of Hilbert completeness, and an operative technology that enabled many remarkable technical achievements, celebrated mostly by Japanese engineers.

It was in the context of this new fuzzy logic, that Vilem Novak carried out what is probably the first work really concerned with the nuts and bolts of natural language semantics from the point of view of fuzzy logic (Novak 1992, 1991). Another notable line of work contributing to a model of fuzzy natural language semantics is that of Ingo Glöckner (Glöckner 2004, 2003, 2001, 2000*c*,*a*,*b*, 1999, 1997*a*,*b*,*c*) concerned with fuzzy natural language quantifiers.

But, despite these successes it may be fair to say that, to this day, fuzzy logic has failed to live up to the high expectations artificial intelligence enthusiasts once had, when they set out to deploy the technology to make machines *understand* the categories of reasoning that humans use to successfully communicate to each other vague ideas and concepts.

Only recently, Zadeh took up renewed interest in this line of research, addressing the main shortcoming when he observes that "progress [in AI] has been, and continues to be, slow in those areas where a methodology is needed in which the objects of computation are perceptions – perceptions of time, distance, form, direction, color, shape, truth, likelihood, intent, and other attributes of physical and mental objects" (Zadeh 2001). The key point Zadeh has to make about perceptions is that they are inherently fuzzy, and that humans use natural language representations where machines use numeric measurements. Thus, the paradigm shift that takes Zadeh into his "new direction of artificial intelligence" is one that takes us "from computing with numbers to computing with words" (Zadeh 1999). The representations of fuzzy concepts employed in his computational theory of perceptions are linguistic in nature. They are expressions of a language he refers to as precisiated natural language (Zadeh 2004*b*). Such a language would have to be natural, in the sense that it is a formal language weakly equivalent to a subset of a natural language, and precisiated, in the sense that every such expression can automatically be translated to a form suitable for approximate reasoning.

Before we turn to some of the more questionable assumptions underlying the visionary end of Zadeh's work and the philosophy of traditional 70-style AI, we have to stress that we do agree with Zadeh in that a technology as envisioned by him, a technology that enables the computational manipulation of linguistic expressions, is highly desirable. In fact we believe that the technology described in this paper is of exactly that nature, and we can think at least of two immediate applications: Natural language interfaces to flexible query processing systems (Zadeh 2003, 2004*a*, Dvorak & Novak 2000), and software tools supporting the implementation of fuzzy controllers in a linguistically intuitive way (Novak 1995, 1997, Bodenhofer & Bauer 2003).

On the other hand, we disagree with Zadeh concerning his silent assumption that a reduction from the problem of computing with words to the strong AI problem is straightforward. For example, Zadeh often cites applications such as parking a car, driving in city traffic, playing golf, or cooking a meal (Zadeh 2001) – those problems that actually do involve perceptions of time, distance, form, direction, color, or shape, and not just perceptions of language as such. His approach therefore presumes that representations of such perceptions in natural languages such as English or German do justice to the actual objects of cognition, which assumes a flavour of Whorfianism possibly too strong for most contemporary minds to savour.

A.2 Degree-based intuitive reasoning

This question of whether or not fuzzy logic is adequate as a model for human cognitive reasoning about vague concepts was taken up in the early 1980s by psychologist Daniel Osherson (Osherson & Smith 1981), who argued that triangular norms cannot account for the intersective concept of a *striped apple*. At this point we have to ask the reader to imagine an *apple* A, and a *striped apple* A'. There can be no doubt that A' is more prototypical of a *striped apple* than of an *apple*, which would have to look like A. This contradicts the non-decreasingness requirement of a t-norm in fuzzy logic and lead, together with a number of other considerations that will not be covered herein, Osherson to conclude that the theory of fuzzy sets does not provide an adequate treatment of a theory of prototypes.

At this point, of course, it may be pointed out that with a *striped apple* more is going on in the semantic domain than a simple intersection. Zadeh also points this out in response to Osherson (Zadeh 1982*a*), maintaining that Osherson's argument doesn't disqualify the theory of fuzzy sets as a basis for a theory of prototypes, but that the major problems lie within prototype theory as such.

In response to this, Osherson et. al. carried out an experiment with naive subjects, asking them to rate the degree of prototypicality of pictured objects like *red apples*, *brown apples*, and *upright apples* (Osherson & Smith 1982). The results were found to be widely inconsistent with any triangular norm. Subjects found the objects more typical of the conjunction than of any of the conjuncts. This is mathematically nonsensical not only for advocates of fuzzy logic. The Kolmogorov axioms, for example, require the same non-decreasingness assumed by fuzzy logic also for probabilities of co-occurences of events, and it turns out that probability theory is equally at odds with naive intuition. This *conjunction fallacy* is now a widely publicized phenomenon that continues to intrigue cognitive psychologists (Sides et al. 2002, Bonini et al. 2004, Tentori et al. 2004, Hertwig & Gigerenzer 1999, Zizzo et al. 2000). As a result it has even been suggested (Huttenlocher & Hedges 1994) that *combined categories* ought to be modeled by bivariate probability distributions, rather than any predefined logical operator. This approach, of course, is highly questionable from a linguistic point of view, as it completely preempts any compositional treatment, so the problem remains. If something as simple and as mathematically fundamental as the non-decreasingness of a conjunction is at odds with human intuition, where can one possibly hope to start in the formalization of an intuitive logic?

In our opinion it is this point where the 70s-style AI tradition has to be abandoned: Fuzzy logic may be successful in applications involving conjunctive concepts, but this does not imply that human intuitive reasoning operates similarly, nor is it fruitful for applications to be modelled after naive intuition. Further evidence in support of this is presented by the functional neuroanatomy of the brain, which seems to separate deductive and probabilistic reasoning (Parsons & Osherson 2001).

A.3 Vagueness

An approach that seems much more fruitful for application-oriented research in fuzzy semantics is to turn away from vague concepts and naive intuitive reasoning to language as such. Apparently vagueness has always been an issue for language, so after Montague semantics (Montague 1973) first offered a systematic model for language meaning in the mid 70s, it did not take long for a degree-based Montague-style semantics to enter the scene (Cresswell 1977). This line of work was eventually refined up until Bierwisch (1989), whose work is still the basis of modern treatments of vagueness. It uses a lambda-calculus to model the way a grammar matches up standards of comparison with vague adjectives, providing an account of comparatives and ordinals in relation to vague predicates. We argue that this work is complementary of, rather than contradictory to, the work in fuzzy logic. Where the linguistic theory of vagueness provides a detailed account of how decision boundaries work in a grammar, fuzzy logic provides a more adequate model of what exactly these decision boundaries look like in the presence of vagueness, than the crisp ones traditionally employed by linguists.

Nevertheless, fuzzy logic never really was adopted into the linguistic body of work on vagueness. One often-cited critic of fuzzy logic in the linguistic community is Manfred Pinkal (Pinkal 1985, 1995). In the context of his work, one has to emphasize that he rejects fuzzy logic not as an alternative to bivalent logic, but rather as an alternative to a logic of vagueness of his own design, which never really took off either.

He argues that fuzzy logic does not resolve the Sorites paradox, predominantly along the following lines of reasoning:

$$\begin{aligned} \mu_{\text{bald}'}(x.\text{hair} = 0) &= 1 \\ \mu_{\text{bald}'}(x.\text{hair} = 1000000) &= 0 \\ \mu_{\text{bald}'}(x.\text{hair} = h) \geq 1 - \epsilon &\Rightarrow \mu_{\text{bald}'}(x.\text{hair} = h + 1) \geq 1 - \epsilon \end{aligned}$$

In words: *If a man with h hair is bald, then so is a man with* $h + 1$ *hair.* This statement is *almost* true (i.e. ϵ is very small, but greater than zero), thus if the antecedent is almost true, so is the consequent. As a result, the paradox still holds.

Here it is important to point out that notions like *almost truth* are not formally part of fuzzy logic, although they are often used in popular expositions. In the fuzzy semantics defined herein, we take a degree of fulfillment $\mu_A(x_1)$ of a proposition $\widetilde{A}$ about x_1 to be meaningful only when compared to the degree of fulfillment $\mu_A(x_2)$ of $\widetilde{A}$ about some x_2. We do not employ it on a meta-level, as Pinkal does. Recall that our setup of the Sorites looks like this:

$$\begin{aligned} \mu_{\widetilde{\text{bald}}'}(x.\text{hair} = 0) &= 1 \\ \mu_{\widetilde{\text{bald}}'}(x.\text{hair} = 1000000) &= 0 \\ \mu_{\widetilde{\text{bald}}'}(x.\text{hair} = h) &\geq \mu_{\widetilde{\text{bald}}'}(h.\text{hair} = h + 1) \end{aligned}$$

Other than modelling the form of a decision boundary as determined by a fuzzy set, and the placement of a decision boundary with respect to a standard of comparison as determined by a grammar, a language model involving fuzzy semantics naturally needs data about boundary placements speakers consider as natural. This empirical line of research was taken up only recently by Kees van Deemter et al. (van Deemter 2006*a*, van Deemter et al. 2006*a*,*b*, van Deemter 2006*b*) in the context of his work on natural language generation involving gradable expressions, and by Moxey and Sanford (Moxey & Sanford 2000, 1997) who concentrate on quantifiers.

Appendix B

Proofs

B.1 Proofs to section 2.1

Theorem 1. *If A is the partition based semantics of a concept on domain X (i.e. $A \subseteq X$), and B is an equivalent representation in terms of ordering based semantics as in definition 3, then B will in fact be a weak ordering on domain X.*

Proof. Clearly by (2.4) we automatically get reflexivity (2.1).

There are four ways to choose x and y from X:

- $x \in X \wedge y \in X$: by (2.5) we have both $(x, y) \in B$ and $(y, x) \in B$.
- $x \in X \wedge y \notin X$: by (2.6) we have both $(x, y) \in B$.
- $x \notin X \wedge y \in X$: by (2.6) we have both $(y, x) \in B$.
- $x \notin X \wedge y \notin X$: by (2.5) we have both $(x, y) \in B$ and $(y, x) \in B$.

Consequently we get completeness (2.3).

Assume $(x, y) \in B$ and $(y, z) \in B$. Here x, y, and z can be chosen in the following ways:

- $x \in A \wedge y \in A \wedge z \in A$: by (2.5) we have $(x, z) \in B$.
- $x \in A \wedge y \in A \wedge z \notin A$: by (2.6) we have $(x, z) \in B$.
- $x \in A \wedge y \notin A \wedge z \in A$: by (2.5) we have $(x, z) \in B$.
- $x \in A \wedge y \notin A \wedge z \notin A$: by (2.6) we have $(x, z) \in B$.
- $x \notin A \wedge y \in A \wedge z \in A$: by (2.6) we have $(x, y) \notin B$, which contradicts our choice of x and y.
- $x \notin A \wedge y \in A \wedge z \notin A$: by (2.6) we have $(x, y) \notin B$, which contradicts our choice of x and y.
- $x \notin A \wedge y \notin A \wedge z \in A$: by (2.6) we have $(y, z) \notin B$, which contradicts our choice of y and z.

- $x \notin A \wedge y \notin A \wedge z \notin A$: by (2.5) we have $(x, z) \in B$.

Consequently we get transitivity (2.2).

□

B.2 Proofs to section 2.2

Theorem 2. *B is the partition based semantics of a concept on domain X iff $A \in \mathcal{P}(X)$, i.e. there is a characteristic function $\chi_A : X \mapsto \{0, 1\}$ where $\chi_A(x) = 1$ iff $x \in B$ and $\chi_A(x) = 0$ iff $x \notin B$. We call A a* crisp *set.*

Proof. Trivial. □

Lemma 1. *If B is the ordering based semantics of a concept on domain X, then there is a characteristic function $\mu_{\widetilde{A}} : X \mapsto [0, 1]$ that ranges over the whole unit interval where $\mu_A(x) \geq \mu_A(y)$ iff $(x, y) \in B$.*

Proof. Let $X = \{x_1, x_2, \ldots, x_n\}$. We can partition X into four sets:

$$\begin{aligned}
&\{x_1\},\\
&L = \{x : (x, x_1) \in B \wedge (x_1, x) \notin B\},\\
&M = \{x \neq x_1 : (x, x_1) \in B \wedge (x_1, x) \in B\},\\
&R = \{x : (x_1, x) \in B \wedge (x, x_1) \notin B\}.
\end{aligned}$$

We can show that L, M, R, and $\{x_1\}$ are disjunct as follows:

- $L \cap M = \emptyset$ since we cannot have both $(x_1, x) \notin B$ as required in the definition of L, and $(x_1, x) \in B$ as required in the definition of M.

- $L \cap R = \emptyset$ since we cannot have both $(x, x_1) \in B$ as required in the definition of L and $(x, x_1) \notin B$ as required in the definition of R.

- $M \cap R = \emptyset$ since we cannot have both $(x, x_1) \in B$ as required in the definition of M and $(x, x_1) \notin B$ as required by the definition of R.

- $x_1 \notin L$ and $x_1 \notin R$, since by letting $x = x_1$ in the definition of L or R respectively, we would have $(x_1, x_1) \in B$ and $(x_1, x_1) \notin B$ which is a contradiction. Furthermore $x_1 \notin M$ by definition of M.

We'll now show that $L \cup M \cup R \cup \{x_1\} = X$. Pick an $x \in X$ arbitrarily. If $x = x_1$ then $x \in \{x_1\}$, so it is in the union. If $x \neq x_1$ there are four cases.

- If $(x, x_1) \in B \wedge (x_1, x) \in B$, then by definition $x \in M$.

- If $(x, x_1) \notin B \wedge (x_1, x) \in B$, then by definition $x \in R$.

- If $(x, x_1) \in B \wedge (x_1, x) \notin B$, then by definition $x \in L$.

- If we had any x with $(x, x_1) \notin B \wedge (x_1, x) \notin B$, then this would contradict (2.3).

We conclude that L, M, R, and $\{x_1\}$ form a partition over X and that L, M, and R have cardinality smaller than $|X|$.

Now let

$$\mu_{\widetilde{A}}^{(X)}(x) = |\{y \in X : \ (x,y) \in B\}|.$$

We will now show in four steps that, equivalently, we can partition X as above and rewrite $\mu_{\widetilde{A}}^{(X)}(x)$ as

$$\mu_{\widetilde{A}}^{(X)}(x) = \begin{cases} \mu_{\widetilde{A}}^{(L)}(x) + 1 + |M| + |R|, & \text{if } x \in L \\ 1 + |M| + |R|, & \text{if } x \in \{x_1\} \\ 1 + \mu_{\widetilde{A}}^{(M)}(x) + |R|, & \text{if } x \in M \\ \mu_{\widetilde{A}}^{(R)}(x), & \text{if } x \in R. \end{cases}$$

First, we can establish that for all $l \in L$ we have $\mu_{\widetilde{A}}^{(X)}(l) = \mu_{\widetilde{A}}^{(L)}(l) + |M| + 1 + |R|$. This is due to the fact that for all $l \in L$

- $l \in X \wedge (l, l') \in B$ for $l' \in L$ iff $l \in L \wedge (l, l') \in B$. (trivially by choice of l)
- $x_1 \in X \wedge (l, x_1) \in B$ by definition of L.
- $m \in X \wedge (l, m) \in B$ for all $m \in M$ since we chose L in such a way that (l, x_1) and M in such a way that (x_1, m). By (2.2) we therefore have $(l, m) \in B$.
- $r \in X \wedge (l, r) \in B$ for all $r \in R$ since by definition of L we have $(l, x_1) \in B$ and by definition of R we have $(x_1, r) \in B$. By (2.2) we therefore have $(l, r) \in B$.

Second, for $x_1 \in \{x_1\}$ we have $\mu_{\widetilde{A}}^{(X)}(x_1) = 1 + |M| + |R|$, because

- $l \in X \wedge (x_1, l) \in B$ for no $l \in L$ by definition of L.
- $x_1 \in X \wedge (x_1, x_1) \in B$ by (2.1). So $\mu_{\widetilde{A}}^{(\{x_1\})}(x_1) = 1$
- $m \in X \wedge (x_1, m) \in B$ for all $m \in M$. by definition of M.
- $r \in X \wedge (m, r) \in B$ for all $r \in R$ by definition of R.

Third, for all $m \in M$ we have $\mu_{\widetilde{A}}^{(X)}(m) = 1 + \mu_{\widetilde{A}}^{(M)}(x) + |R|$, because

- $l \in X \wedge (m, l) \in B$ for no $l \in L$. Assume for the sake of contradiction that there was such an l. By definition of M we would have $(x_1, m) \in B$. By (2.2) we would have $(x_1, l) \in B$, which contradicts the definition of L.
- $x_1 \in X \wedge (m, x_1) \in B$ by definition of M.
- $m' \in X \wedge (m, m') \in B$ for $m' \in M$ iff $m \in M \wedge (m, m') \in B$. (trivially by choice of m)
- $r \in X \wedge (m, r) \in B$ for all $r \in R$ since by definition of M we have $(m, x_1) \in B$ and by definition of R we have $(x_1, r) \in B$. By (2.2) we therefore have $(m, r) \in B$.

Fourth, for all $r \in R$ we have

- $l \in X \wedge (r, l) \in B$ for no $l \in L$. Assume for the sake of contradiction that there was such an l. By definition of R we would have $(x_1, r) \in B$. By (2.2) we would have $(x_1, l) \in B$, which contradicts the definition of L.
- $x_1 \in X \wedge (r, x_1) \notin B$ by definition of R.
- $m \in X \wedge (r, m) \in B$ for no $m \in M$ since, by definition of M we would have $(m, x_1) \in B$. By (2.2) we would have $(r, x_1) \in B$, which contradicts the definition of R.
- $r \in X \wedge (r', r) \in B$ for $r' \in R$ iff $r \in R \wedge (r', r) \in B$. (trivially by choice of r)

Next, note that, by definition, for all $x \in X$, $\mu^{(L)}_{\widetilde{A}}(x) \leq |L|$, $\mu^{(M)}_{\widetilde{A}}(x) \leq |M|$, $\mu^{(R)}_{\widetilde{A}}(x) \leq |R|$. Consequently, for all x we have

$$\mu^{(L)}_{\widetilde{A}}(x) + 1 + |M| + |R| \quad \geq \quad 1 + |M| + |R| \quad \geq \quad 1 + \mu^{(M)}_{\widetilde{A}}(x) + |R| \quad \geq \quad \mu^{(R)}_{\widetilde{A}}(x).$$

We can now show by induction on $|X|$ that if $\widetilde{A}$ is the ordering based meaning of a fuzzy concept on domain X, then for all $x, y \in X$, $\mu^{(X)}_{\widetilde{A}}(x) \geq \mu^{(X)}_{\widetilde{A}}(y)$ iff $(x, y) \in B$.

Base If $X = \{\}$, then there is no element to choose from X, so Lemma 3 holds vacuously.

Base If $X = \{x_1\}$, then there is one element to choose from X, so for the x and y above we have $x = y = x_1$. By (2.1) we always have $(x, y) \in B$, and of course we always have $\mu_{\widetilde{A}}(x) = \mu_{\widetilde{A}}(y)$.

Induction Let $X = \{x_1, x_2, \ldots, x_n\}$. We can now partition X as above into sets L, M, R. Since all of these have cardinality smaller than $|X|$ we can assume the above statement w.r.t. domains L, M, and R by inductive hypothesis. If both x and y are in either L, M, or R then our recursive definition of $\mu^{(X)}_{\widetilde{A}}(x)$ applies the same case to them, thus adding the same constant to $\mu^{(\cdot)}_{\widetilde{A}}(x)$ and $\mu^{(\cdot)}_{\widetilde{A}}(y)$, which preserves the ordering. Otherwise there are twelve ways to distribute the choices of x and y. Since, by 2.3, $(x, y) \in B$ or $(y, x) \in B$ we can assume $(x, y) \in B$ without loss of generality which leaves six ways to distribute these choices that are consistent with the definitions of L, M, and R. In each of these cases $\mu^{(\cdot)}_{\widetilde{X}}(x) \geq \mu^{(\cdot)}_{\widetilde{X}}(y)$.

1. For $x \in L$, $y = x_1$ note $\mu^{(L)}_{\widetilde{A}}(x) + 1 + |M| + |R| \quad \geq \quad 1 + |M| + |R|$.
2. For $x \in L$, $y \in M$ note $\mu^{(L)}_{\widetilde{A}}(x) + 1 + |M| + |R| \quad \geq \quad 1 + \mu^{(M)}_{\widetilde{A}}(x) + |R|$.
3. For $x \in L$, $y \in R$ $\mu^{(L)}_{\widetilde{A}}(x) + 1 + |M| + |R| \quad \geq \quad \mu^{(R)}_{\widetilde{A}}(x)$.
4. For $x = x_1$, $y \in M$ $1 + |M| + |R| \quad \geq \quad 1 + \mu^{(M)}_{\widetilde{A}}(x) + |R|$.
5. For $x = x_1$, $y \in R$ $1 + |M| + |R| \quad \geq \quad \mu^{(R)}_{\widetilde{A}}(x)$
6. For $x \in M$, $y \in R$ $1 + \mu^{(M)}_{\widetilde{A}}(x) + |R| \quad \geq \quad \mu^{(R)}_{\widetilde{A}}(x)$

Thereby our choice for $\mu_{\widetilde{A}}^{(X)}(x)$ would constructively prove lemma 1, if it wasn't for the fact that lemma 1 requires a characteristic function $\mu_{\widetilde{A}}(x)$ that ranges over the unit interval. However, this is easily established by letting

$$\mu_{\widetilde{A}}(x) = \frac{\mu_{\widetilde{A}}^{(X)}(x)}{|X|},$$

which completes the proof. □

Lemma 2. *If there is a characteristic function $\mu_{\widetilde{A}} : X \mapsto [0, 1]$ that ranges over the whole unit interval where $\mu_A(x) \geq \mu_A(y)$ iff $(x, y) \in B$, then B is the ordering based meaning of a concept on domain X,*

Proof. We can trivially let

$$B = \{(x, y) : \ \mu_A(x) \geq \mu_A(y)\}.$$

Since $\geq$ is a partial ordering on the unit interval, any function with the unit interval as the codomain imposes at least a weak ordering on its domain. □

Theorem 3. *B is the ordering based semantics of a concept on domain X iff $\widetilde{A} \in \widetilde{\mathcal{P}}(X)$, i.e. there is a characteristic function $\mu_{\widetilde{A}} : X \mapsto [0, 1]$ that ranges over the whole unit interval where $\mu_{\widetilde{A}}(x) \geq \mu_{\widetilde{A}}(y)$ iff $(x, y) \in B$.*

Proof. Trivial from lemmata 1 and 2. □

Appendix C

Experimental setup

C.1 Linguistic data model

```
LEXENT adv {
  STEM "very";
  TYPE "adv_degree_spec_le";
};

LEXENT adv {
  STEM "quite";
  TYPE "adv_degree_spec_le";
};

LEXENT adv {
  STEM "rather";
  TYPE "adv_degree_spec_le";
};



ENTITY place {

  LEXENT noun {
    STEM "city";
    TYPE "n_intr_le";
    ONSET "con";
  };

  LEXENT noun {
    STEM "town";
    TYPE "n_intr_le";
    ONSET "con";
  };
```

```
GEN nb "#noun";

PK placeid;

ID(100) placename {
  TYPE "n_proper_city_le";
  ONSET "con";
};

REFERENCE refstate TO ONE state;

REFERENCE refnear TO MANY place {
  INTAT distance {

    LEXENT near {
  STEM "near";
  TYPE "p_reg_le";
  ONSET "con";
  REL "_NEAR_P_REL";
};

DSCR "If a city was <B>a distance #distance</B> kilometres from another city, "
  "it would be natural to say that the cities are <B>#near</B> each other.";

  };
};

STRAT(10) type;

INTAT lat;
INTAT long;

INTAT pop {

  LEXENT adj {
    STEM "tiny";
    TYPE "adj_intrans_le";
  };

  LEXENT adj {
    STEM "small";
    TYPE "adj_intrans_le";
  };

  LEXENT adj {
    STEM "big";
```

```
      TYPE "adj_intrans_le";
    };

    LEXENT adj {
      STEM "huge";
      TYPE "adj_intrans_le";
    };

    GEN ap "#adv #adj";
GEN nb "#ap #noun";

    DSCR "If a city had a <B>population of #pop</B>, "
      "it would be natural to call it a <B>#ap</B> place.";

  };

  INTAT temp {

    LEXENT adj {
      STEM "hot";
      TYPE "adj_intrans_le";
      ONSET "con";
    };

    LEXENT adj {
      STEM "cold";
      TYPE "adj_intrans_le";
      ONSET "con";
    };

    GEN ap "#adv #adj";
GEN nb "#ap #noun";

    DSCR "If a city had a year-round average <B>temperature of #temp</B> "
       "degrees celsius, it would be natural to call it a <B>#ap</B> city.";

  };

  INTAT wet {

    LEXENT adj {
      STEM "rainy";
      TYPE "adj_intrans_le";
      ONSET "con";
    };
```

```
    LEXENT adj {
      STEM "dry";
      TYPE "adj_intrans_le";
      ONSET "con";
    };

    GEN ap "#adv #adj";
GEN nb "#ap #noun";

    DSCR "If in a city the weather was <B>rainy #wet</B> days of "
      "the year, it would be natural to call it a <B>#ap</B> city.";

  };

};

ENTITY state {

  PK stateid;
  STRAT(2) state;

};



ENTITY skyscraper {

  LEXENT noun {
    STEM "tower";
    TYPE "n_intr_le";
    ONSET "con";
  };

  LEXENT noun {
    STEM "skyscraper";
    TYPE "n_intr_le";
    ONSET "con";
  };

  LEXENT noun {
    STEM "building";
    TYPE "n_intr_le";
    ONSET "con";
  };

  GEN nb "#noun";
```

```
  PK scraperid;

  ID(200) scrapername {
    TYPE "n_proper_le";
    ONSET "con";
  };

  INTAT floors {

    LEXENT adj {
      STEM "big";
      TYPE "adj_intrans_le";
    };

    LEXENT adj {
      STEM "small";
      TYPE "adj_intrans_le";
    };

    GEN ap "#adv #adj";
GEN nb "#ap #noun";

    DSCR "If a skyscraper had <B>#floors floors</B>, it would be natural to call it "
      "a <B>#ap</B> skyscraper.";

  };

  INTAT completion {

    LEXENT adj {
      STEM "new";
      TYPE "adj_intrans_le";
    };

    LEXENT adj {
      STEM "old";
      TYPE "adj_intrans_le";
    };

    GEN ap "#adv #adj";
GEN nb "#ap #noun";

    DSCR "If a skyscraper had been completed in the <B>year #completion</B>, "
      "it would be natural to call it a <B>#ap</B> skyscraper.";
```

```
  };

};
```

C.2 Instructions

If, one day, we will be able to talk to our computers, we will formulate database queries by statements like 'Computer, list all big cities near San Francisco', or 'Computer, give me all cities with lots of tall new skyscrapers'. But what population exactly does it take for a city to be a *big city*? What distance can it be from San Francisco, so that it's a *big city near San Francisco*?

This experiment is designed to bring us one step closer to that vision of interfacing with databases in natural language by collecting some data about how exactly we use vague expressions like *big*, *near*, *tall*, and so on.

In order to do so we will have to ask you some questions which are generally of the form 'Given the following information about some thing, one can call it XXX.' More particularly, there are three ways the computer will formulate questions.

- When the computer uses a text field, you are invited to make suggestions about what kind of expressions you would intuitively want to use in the context provided. This information serves as an experimental control and will be analyzed by humans, not by the computer, so you can treat it as if you were talking to a human.

- Sometimes the computer will generate a single word, and ask whether the resulting sentence is true or false. These questions are always asked while systematically varying the data you are given.

- Sometimes the computer will randomly generate an expression like *very small* or *rather near* and ask you whether the generated sentence is correct or not.

The different types of questions and series of systematically generated questions will alternate randomly. There is no need for you to try and keep track of what the computer is doing. You need to rely on your ad-hoc intuition about each individual question only. Some questions may also be asked multiple times, due to a limitation in the software.

The experiments are generally about the adjectives, not the nouns! If the system asks you to judge whether a skyscraper with 7 floors would be called a small skyscraper, don't answer *false*, because you think it's not a skyscraper. Simply proceed under the hypothesis presented in the if-part. That is, if a building with 7 floors *was* a skyscraper, it would definitely be a small one.

You will find some questions very straightforward to answer. A 2-metre basketball player is unquestionably tall, a 1.5-metre polo jockey is definitely not tall. But what about your average 1.75-metre male adult football fan? Is he tall? We do realize that it is awkward to call such a person *tall*, just as much as calling them *small.* Nevertheless, you will be asked to make such judgements, and you might have the feeling that you keep changing your mind about those questions. Here we have to ask you to remember that you are not being tested. We realize that people don't use vague expressions consistently all the time. What we are trying to find out is just how people do use them. So don't take too much time on any of the questions. Don't try to memorize and structure your answers, if that is not what you naturally do when talking about tall basketball players and small polo jockeys to your friends in the pub. Try to answer the questions from your natural intuition.

If you feel you don't have good intuitions about the numbers provided, for example the distance experiment displaying numbers in kilometres or the temperature experiment displaying temperatures in degrees celsius when you are unaccustomed to these units, then leave these experiments for last. Run a couple of numbers through a converter to try and get some feeling for them before doing the experiment.

Sometimes it may seem to you that some relevant contextual information is missing. For example for the distance experiment you may ask yourself whether to judge this by European standards or by US standards. In these cases just make an assumption that makes sense for you, and try to stick with this assumption.

The language abilities of our computer system are very limited: Sometimes you may find that sentences are being generated that sound a bit awkward. In those cases, don't judge the sentence to be false just because you wouldn't normally say it yourself. Try to imagine the system talking to you in a strong foreign accent, and judge whether or not you would generally understand what is meant.

About the software This web-application has been designed to make the data-collection process as efficient and as comfortable for you as possible. After entering your username you are asked to make a selection for one particular sub-experiment. The numbers in parentheses show how many questions you have already answered, and approximately how many questions there are to answer. You can select any one of them, and the computer will start asking you some questions. After you are done with one such batch of questions, the computer will redirect you to the page you started from. You can then select another batch of questions. During the question-answering *don't use the back and forward keys of your browser!*

Although we would appreciate it if you could try to do as much of the experiment as possible, we do respect your schedule and the time you are dedicating to this experiment. This is why you can leave the experiment at any point. Even partially completed experiments will provide valuable data for us. If you log in again with the same username, you will be able to continue the experiment just where you left off.

Please do get in touch! Please feel free to email us any time at `rbergmair@acm.org`, if you would like further details on the experiment, if you are interested in the final report of this project, or if you have any questions, ideas, or suggestions. We are looking forward to any feedback from you!

Thanks a lot for taking part in this experiment!

Bibliography

Androutsopoulos, I. (1995), 'Natural language interfaces to databases: an introduction', *Natural language engineering* **1**(1), 29–81.

Androutsopoulos, I. (2000), Database interfaces, *in* R. D. et al., ed., 'Handbook of Natural Language Processing', pp. 209–240.

Bergmair, R. (2006), Syntax-driven analysis of context-free languages with respect to fuzzy relational semantics, Technical Report UCAM-CL-TR-663, University of Cambridge, Computer Laboratory.

Bergmair, R. & Bodenhofer, U. (2006), Syntax-driven analysis of context-free languages with respect to fuzzy relational semantics, *in* 'Proceedings of the 2006 IEEE World Congress on Computational Intelligence (WCCI '2006)'.

Bierwisch, M. (1989), The semantics of gradation, *in* M. Bierwisch & E. Lang, eds, 'Dimensional Adjectives', pp. 71–261.

Bodenhofer, U. & Bauer, P. (2003), A formal model of interpretability of linguistic variables, *in* J. Casillas, O. Cordón, F. Herrera & L. Magdalena, eds, 'Interpretability Issues in Fuzzy Modeling', Vol. 128 of *Studies in Fuzziness and Soft Computing*, Springer-Verlag, Heidelberg, pp. 524–545.

Bonini, N., Tentori, K. & Osherson, D. (2004), 'A different conjunction fallacy', *Mind and Language* **19**(2), 199–210.

Copestake, A. (2004), Robust minimal recursion semantics, unpublished draft.

Copestake, A. & Spärck-Jones, K. (1990), 'Natural language interfaces to databases', *Knowledge Engineering Review* **5**(4), 225–249.

Copestake, A., Flickinger, D., Sag, I. & Pollard, C. (1999), Minimal recursion semantics: An introduction, preprint available from `http://www.cl.cam.ac.uk/~aac10/papers/newmrs.pdf`.

Cresswell, M. (1977), The semantics of degree, *in* B. Partee, ed., 'Montague Grammar', pp. 261–292.

De Cock, M. & Kerre, E. E. (2002), 'A context-based approach to linguistic hedges', *International Journal of Applied Mathematical Computer Science* **12**(3), 371–382.

De Cock, M., Bodenhofer, U. & Kerre, E. E. (2000), Modelling linguistic expressions using fuzzy relations, *in* 'Proceedings of the 6th International Conference on Soft Computing', Iizuka, pp. 353–360.

DELPH-IN (2006), 'Blue delph-in: Open-source deep processing', website. `http://www.delph-in.net/`.

Dvorak, A. & Novak, V. (2000), On the extraction of linguistic knowledge in databases using fuzzy logic, *in* H. L. Larsen, ed., 'Flexible Query Answering Systems. Recent Advances.', Physica-Verlag, pp. 445–454.

Gaines, B. R. & Kohout, L. J. (1977), 'The fuzzy decade: A bibliography of fuzzy systems and closely related topics', *International Journal of Man-Machine Studies* **9**, 1–68.

Glöckner, I. (1997*a*), A framework for evaluating approaches to fuzzy quantification, Technical Report TR97-06, Universität Bielefeld Technische Fakultt, PO-Box 100131, 33501 Bielefeld, Germany.

Glöckner, I. (1997*b*), Fuzzy quantifiers for processing natural-language queries in content-based multimedia retrieval systems, Technical Report TR97-05, Universität Bielefeld Technische Fakultt, PO-Box 100131, 33501 Bielefeld, Germany.

Glöckner, I. (1997*c*), A system for the content-based retrieval of textual and non-textual documents based on natural language queries, Technical Report TR97-03, Universität Bielefeld Technische Fakultt, PO-Box 100131, 33501 Bielefeld, Germany.

Glöckner, I. (1999), A framework for evaluating approaches to fuzzy quantification, Technical Report TR99-03, Universität Bielefeld Technische Fakultt, PO-Box 100131, 33501 Bielefeld, Germany.

Glöckner, I. (2000*a*), Advances in dfs theory, Technical Report TR2000-01, Universität Bielefeld Technische Fakultt, PO-Box 100131, 33501 Bielefeld, Germany.

Glöckner, I. (2000*b*), An axiomatic theory of fuzzy quantifiers in natural languages, Technical Report TR2000-03, Universität Bielefeld Technische Fakultt, PO-Box 100131, 33501 Bielefeld, Germany.

Glöckner, I. (2000*c*), A broad class of standard dfses, Technical Report TR2000-02, Universität Bielefeld Technische Fakultt, PO-Box 100131, 33501 Bielefeld, Germany.

Glöckner, I. (2001), Standard models of fuzzy quantification, Technical Report TR2001-01, Universität Bielefeld Technische Fakultt, PO-Box 100131, 33501 Bielefeld, Germany.

Glöckner, I. (2003), Fundamentals of fuzzy quantification: Plausible models, constructive principles, and efficient implementation, Technical Report TR2002-07, Universität Bielefeld Technische Fakultt, PO-Box 100131, 33501 Bielefeld, Germany.

Glöckner, I. (2004), 'Evaluation of quantified propositions in generalized models of fuzzy quantification', *International Journal of Approximate Reasoning* **37**(2), 93–126.

Goguen, J. A. (1974), 'Concept representation in natural and artificial languages: Axioms, extensions and applications for fuzzy sets', *International Journal of Man-Machine Studies* **6**, 513–561.

Goguen, J. A. (1975), On fuzzy robot planning, *in* L. A. Zadeh, ed., 'Fuzzy Sets and their Applications to Cognitive and Decision Processes', Academic Press.

Hertwig, R. & Gigerenzer, G. (1999), 'The 'conjunction fallacy' revisited: How intelligent inferences look like reasoning errors', *Journal of Behavioral Decision Making* **12**, 275–305.

Huttenlocher, J. & Hedges, L. V. (1994), 'Combining graded categories: Membership and typicality', *Psychological Review* **101**(1), 157–165.

IPCC (2005), 'Climate baseline data: 10-year means', website. available from `http://ipcc-ddc.cru.uea.ac.uk/obs/get_10yr_means.html`.

Klement, E. P., Mesiar, R. & Pap, E. (2000), *Triangular Norms*, Vol. 8 of *Trends in Logic*, Kluwer Academic Publishers, Dordrecht.

Lee, E. T. & Zadeh, L. A. (1969), 'Note on fuzzy languages', *Information Sciences* **1**, 421–434.

Montague, R. M. (1973), The proper treatment of quantification in ordinary english, *in* K. J. J. Hintikka & J. M. E. Moravcsik, eds, 'Approaches to Natural Language'.

Moxey, L. & Sanford, A. (1997), Choosing the right quantifier, *in* T. Givon, ed., 'Conversation: cultural, communicative and social perspectives', pp. 207–231.

Moxey, L. & Sanford, A. (2000), 'Communicating quantities: A review of psycholinguistic evidence of how expressions determine perspective', *Applied Cognitive Psychology* **14**, 237–255.

Novak, V. (1991), 'Fuzzy logic, fuzzy sets, and natural languages', *International Journal on General Systems* **20**, 83–97.

Novak, V. (1992), *The Alternative Mathematical Model of Linguistic Semantics and Pragmatics*, Vol. 8 of *IFSR International Series on Systems Science and Engineering*, International Federation for Systems Research.

Novak, V. (1995), 'Linguistically oriented fuzzy logic controller and its design', *International Journal of Approximate Reasoning*.

Novak, V. (1997), Evaluating linguistic expressions and their role in the design of the fuzzy control strategy, *in* N. Steele, ed., 'Proceedings of the Second International ICSC Symposium on Fuzzy Logic and Applications (ISFL '97)', ICSC, ICSC Academic Press, pp. 89–94.

Osherson, D. & Smith, E. (1981), 'On the adequacy of prototype theory as a theory of concepts', *Cognition* **9**, 35–58.

Osherson, D. & Smith, E. (1982), 'Gradedness and conceptual combination', *Cognition* **12**, 299–318.

Parret, H. (1974), *Discussing Language*, Vol. 93 of *Janua Linguarum Series Maior*, Mouton.

Parsons, L. & Osherson, D. (2001), 'New evidence for distinct right and left brain systems for deductive vs. probabilistic reasoning', *Cerebral Cortex* **11**(10), 951–965.

Pinkal, M. (1985), *Logik und Lexikon*, Berlin.

Pinkal, M. (1995), *Logic and Lexicon*, Kluwer, Dordrecht.

Sheppard, D. (1954), 'The adequacy of everyday quantitative expressions as measurements of qualities', *British Journal of Psychology* **45**, 40–50.

Sides, A., Osherson, D., Bonini, N. & Viale, R. (2002), 'On the reality of the conjunction fallacy', *Memory & Cognition* **30**, 191–198.

Tentori, K., Bonini, N. & Osherson, D. (2004), 'The conjunction fallacy: a misunderstanding about conjunction?', *Cognitive Science* **28**, 467–477.

USCB (2000), 'U.s. gazetteer 2000', website. available from `http://www.census.gov/geo/www/gazetteer/gazette.html`.

van Deemter, K. (2006*a*), 'Generating referring expressions that involve gradable properties', *Computational Linguistics*. to appear.

van Deemter, K. (2006*b*), 'Informing generation algorithms through experiments with human subjects: the case of gradable descriptions', *Computational Linguistics*. submitted.

van Deemter, K., van der Sluis, I. & Gatt, A. (2006*a*), Building a semantically transparent corpus for the generation of referring expressions, *in* 'Proceedings of the 4th International Natural Language Generation Conference (INLG '06)'.

van Deemter, K., van der Sluis, I. & Gatt, A. (2006*b*), Building a semantically transparent corpus for the generation of referring expressions, *in* 'Proceedings of the 4th International Natural Language Generation Conference (INLG '06)'.

Zadeh, L. A. (1965), 'Fuzzy sets', *Information and Control* **8**, 338–353.

Zadeh, L. A. (1975*a*), 'The concept of a linguistic variable and its application to approximate reasoning Part I', *Information Sciences* **8**, 199–250.

Zadeh, L. A. (1975*b*), 'The concept of a linguistic variable and its application to approximate reasoning Part II', *Information Sciences* **8**, 301–357.

Zadeh, L. A. (1975*c*), 'The concept of a linguistic variable and its application to approximate reasoning Part III', *Information Sciences* **9**, 43–80.

Zadeh, L. A. (1978), 'Pruf – a meaning representation language for natural languages', *International Journal of Man-Machine Studies* **10**, 395–460.

Zadeh, L. A. (1981), Test-score semantics for natural languages and meaning representation via pruf, *in* B. B. Rieger, ed., 'Empirical Semantics: A collection of new approaches in the field.', Studienverlag Brockmeyer, pp. 281–349.

Zadeh, L. A. (1982*a*), 'A note on prototype theory and fuzzy sets', *Cognition* **12**, 291–297.

Zadeh, L. A. (1982*b*), Test-score semantics for natural languages, *in* 'Proceedings of the 9th conference on Computational linguistics', Academia Praha, Czechoslovakia, pp. 425–430.

Zadeh, L. A. (1999), 'From computing with numbers to computing with words. from manipulation of measurements to manipulation of perceptions', *IEEE Trans. Circuits Syst. I* **46**, 105–119.

Zadeh, L. A. (2001), 'A new direction in ai: Toward a computational theory of perceptions', *AI Magazine* **22**, 73–84.

Zadeh, L. A. (2003), From search engines to question-answering systems – the need for new tools, *in* 'Web Intelligence, First International Atlantic Web Intelligence Conference (AWIC 2003)', Vol. 2663 of *Lecture Notes in Computer Science*, Springer, pp. 15–17.

Zadeh, L. A. (2004*a*), 'A note on web intelligence, world knowledge, and fuzzy logic', *Data & Knowledge Engineering* **50**, 291–304.

Zadeh, L. A. (2004*b*), 'Precisiated natural language', *AI Magazine*.

Zizzo, D., Stolarz-Fantino, S., Wen, J. & Fantino, E. (2000), 'A violation of the monotonicity axiom: the experimental evidence on the conjunction fallacy', *Journal of Economic Behavior and Organization* **41**, 263–276.