Lexical Meaning and Disambiguation

Henk Zeevat
SFB991, Heinrich Heine Universitaet Duesseldorf &
ILLC, University of Amsterdam

Sofia Ambiguity Workshop

An argument

Ambiguity is theoretically badly understood.
Linguists who are not computational have no idea of the importance of ambiguity

Ergo: this is a very important workshop

Apology

Giving a talk on stochastic frames in which ambiguity was an example, people found the application puzzling. This is an attempt at giving a new presentation of old joint work with Scott Grimm, Lotte Hogeweg, Sander Lestrade and Elisabeth Allyn-Smith. We tried to convert linguists, maybe it is also a good idea to explain this view to people who are convinced that ambiguity is important.

MIU (meaning in use)

an answer to the question what a word w expresses in a concrete use

assumption I

MIUs can be equated with frames in a suitable frame semantics

assumption II

MIU frames can be represented by a finite number of binary semantic features

Evidence

templates from PATR 2 where the frame is the unification of the templates

a decomposition of the frames

path1 = path2
path:value
path

Ambiguity

A word in its use typically has many MIUs

A theory of (lexical) ambiguity

A. How to construct the MIU for a word that applies in the context?

B. What is the right word to use given that you want the hearer to understand it as a particular MIU

C. How is the lexical knowledge structured to allow fast decisions on these two questions?

D. How can that structure be acquired?

D. How can that structure be acquired?

By experience.

Our linguistic experience contains words with their MIUs in different contexts.

$E(G) = W$ where W is the set of words that were used for MIU G.

with a preference ordering: how often this happened. So W should have elements (w, n) or $(w, n/|W|)$

That gives a good but incomplete answer to B

Dictionary view of ambiguity

Invert E

$meanings(w) = \{G : \exists r(w, r) \in E(G)\}$

gives many meanings

can acquire frequencies: how often does w express G?

best resolution strategy: take the most frequent G

WSD: try to improve on that using the context

Smolensky's and Hogeweg's account of ambiguity

coffee example

coffee associates with many microfeatures, associated to the word by learning

beans
powder
liquid
brown
hot
pungent
bushes
red

the context switches off many of these, e.g. in cup of coffee

x beans
x powder
liquid
brown
hot
pungent
x bushes
x red

Hogeweg

Optimality Theory

fit $>$ strength

Strength: project associated features

Fit: but not those that are inconsistent with the context

Good performance on Dutch particle wel and other highly grammaticalised words, less obvious on lexical words

Like other interpretational Optimality Theory, it seems a reformulation of Bayesian Interpetation:

the prefered reading maximises the product of the prior and likelihood

prior: Fit

likelihood: each extra feature increases the probability that the word is used

That makes the OT system true in an approximative sense. Low priors may be important, not all associated features need to contribute to likelihood

Problem:

flat sets of features

obligatory and optional features

implicational connections between features

forced choice

defaults

Zeevat, Grimm, Hogeweg, Lestrade, Allyn Smith

replace sets of features by abstract joint distributions over features

joint distribution for a word:

$p_w(G)$ where G is a set of features

abstract:

$p = q$ iff $\forall xy\ p(x) < p(y) \leftrightarrow q(x) < q(y)$

1. one learns an abstract distribution by learning a distribution

2. abstract distributions can be described by stochastic frames

3. the stochastic frames can be "resolved" by adding contextual information to give a concrete frame, a meaning in use.

1. needs no further explanation. It is just one way to represent a corpus of learning data.

2. requires the insight that constraints can be written as probabilistic (in)equalities

For the application, the following are needed.

absolute features: $p(F) = 1$

Other features are defaults only: they are there if they can be, but can be switched off. But there can be preferences:

$p_w(F) > p_w(G)$

conditionally absolute features: $p(F|G) = 1$

Feature F can become absolute if G is assumed.

notation $G \Rightarrow F$

features that exclude each other: $p(F \wedge G) = 0$

forced choice between a finite set of features:
$p(F1 \vee \ldots \vee Fn) = 1$ and $p(Fi \wedge Fj) = 0$ for $i < j \leq n$

In a forced choice, there can be preferences and normally there are.

Notation

$x \in \{y_1, \ldots, y_n\}$

x is forced to take the first compatible value from $y_1 \ldots y_n$.

Case study: the verb fall

A stochastic frame is presented that collects 64 attested truth conditionally non equivalent readings of the verb fall taken from

Russian, English, Dutch, French and German, using dictionaries and the internet.

THEME $\in$ {concrete, light, precipitation, task, date, judgment, proposal}

A forced choice for the THEME (the theme of the verb) to take a value in the indicated classes.

nocontrol(THEME)
nocause(THEME)

Makes the theme a non-agent.

SOURCE: location(DIMENSION)
POSITION: location(DIMENSION)
DIMENSION $\in$ {space, posture, life, health, moral, quantity, level, outcome(PROCESS)}

Introduces a source, a destination (POSITION) and a dimension in which the movement takes place.

SOURCE= PART1
split(PART1, PART2, THEME)

Her hair fell perfectly. This forces a split of the theme into top and bottom.

at(POSITION, down(SOURCE,DIMENSION))

The destination is below the source in the dimension.

The above information is presupposed (from the perspective of the verb). Now follows the message:

EVENT: movement(THEME,SOURCE,POSITION)

the event is a movement of the theme from source to destination.

at(THEME,POSITION)

If PART2 replaces THEME:
at(PART2, POSITION)

This is the result:

Given:

THEME $\in$ { concrete, light, precipitation, task, date, judgment, proposal} Absolute
nocause(THEME)
nocontrol(THEME) Absolute
SOURCE: location(DIMENSION)
POSITION: location(DIMENSION)
DIMENSION $\in$ {space (default), posture, life, health, moral, quantity, level, outcome(PROCESS)} Absolute
SOURCE= PART1 $\Leftarrow$ split(PART1,PART2, THEME) Absolute
split(PART1, PART2, THEME)
at(POSITION, down(SOURCE,DIMENSION)) Absolute

New:

at(THEME,POSITION) Default
at(PART2, POSITION) $\Leftarrow$ split(PART1,PART2, THEME)
Absolute
EVENT:movement(THEME,SOURCE) $\Leftarrow$ at(THEME,
SOURCE) Absolute

The above is not frame notation, but can be reformulated as such without complications at the cost of readability.

This is a little theory about variables at the end of frame paths.

It is a stochastic frame, by interpreting the extra annotations ($\in$, $\Leftarrow$, Absolute) as indicated. It is a frame with its absolute part given as usual and the rest of its structure given by a set of joint distributions.

The context (e.g. the NP binding the THEME or the PP binding the SOURCE) can disambiguate the frame by putting non-absolute features to 0 or 1.

The following gives some examples.

The glass fell on the floor.

THEME:glass
glass:concrete
nocontrol(glass)
nocause(glass)
SOURCE:location(space)
at(floor,down(SOURCE,space))
at(glass,floor)
movement(glass,SOURCE)

John fell.

THEME:John
John:concrete
nocontrol(John)
nocause(John)
SOURCE: standing
POSITION: lying down
lying down:location(posture)
at(lying down,down(standing,posture))
at(John,lying down)
movement(John,standing)

The house fell.

THEME:house
house:concrete
nocontrol(house)
nocause(house)
SOURCE:erect
POSITION:collapsed
collapsed:location(posture)
at(collapsed,down(erect,posture))
at(house,collapsed)
movement(house,erect)

Special constraint needed:

if the dimension is life, the theme is a soldier and the movement is part of a battle

The corporal fell.

THEME:corporal
corporal:concrete
nocontrol(corporal)
nocause(corporal)
SOURCE:alive
POSITION:dead
dead:location(life)
at(dead,down(alive,life))
at(corporal,dead)
movement(corporal,alive)

The rain fell.

THEME:rain
rain:precipitation
nocontrol(rain)
nocause(rain)
SOURCE:sky
POSITION:ground
ground:location(space)
at(ground,down(sky,space))
at(rain,ground)
movement(rain,sky)

Christmas day falls on a Sunday this year.

THEME:Christmas
Christmas: date
nocontrol(Christmas)
nocause(Christmas)
SOURCE: location(outcome(calendar))
POSITION: Sunday
at(Sunday,down(calendar,outcome(calendar)))
at(Christmas, Sunday)

De browser valt na 5 minuten uit.
The browser falls after 5 minutes out.
The browser crashes after 5 minutes.

PROCESS: browser
THEME:browser
nocontrol(browser)
nocause(browser)
SOURCE: functioning
POSITION: non-functioning
non-functioning:location(outcome(browser))
at(non-functioning,down(functioning,outcome(browser))
at(browser, non-functioning)
movement(browser,functioning,non-functioning)

De taak valt op mij.
the task falls on me
It is my task.

nocontrol(task)
nocause(task)
ME:location(outcome(taskassignment))
at(ME,down(taskassignment,outcome(taskassignment))
at(task,ME)

Some conclusions from fall

1. It has been tested by software and seems to work for all occurrences of fall in Martin Chuzzlewith (Jonathan Mallinson)

One problem: sometimes it is not the sentence that gives the needed contextual information.

The context gives enough information to do the disambiguation and this may be a constraint on new uses in the historical development of the word.

2. differences between languages are minimal and mostly transparent

Grief fell from his heart.
Schmerz fiel von seinem Herzen
Er fiel in die Sachsen.
He fell onto the Saxons
De prijs valt op mij
The prize falls on me.

3. not necessary that the same uses are idiomatic

Tolstoy (a major source for Russian dictionaries) tends to import German idiom but has not succeeded in making that Russian idiom.

Obviously his uses are understood and appreciated.

Typically we do not need instruction in the idiomatic uses in other languages (except marginally)

Second case study: run, walk, laufen, rennen

Observation

run and rennen
walk and laufen

no significant differences in their primary uses

but very significant differences in their secondary uses

* Der Nase/Die Maschine/Die Dusche rennt.
His nose/the machine/the tap is running.
Der Nase/Die Maschine/Die Dusche lauft.
*His nose/the machine/the tap is walking.

The uses are obviously metaphoric extensions.

Prediction:
while starred, the uses are quite understandable.

Problem 1

Explain the stars

Problem 2

Explain production

Corpus

a function E from $pow(FEAT)$ to sets of word-probability pairs

$(w, r) \in E(G)$ iff G is the precise contribution of w at some point in the corpus and this happens n times and m is the number of times G is the exact contribution of some word w and $r = n/m$

Attempted production rule

To express G in c use w if $(w, r) \in F(G)$

Effect

*Der Nase rennt.
etc.

Problem: G may well not be recoverable by the hearer

Solution:

Make the production rule dependent on recoverability in c.

Second problem

No innovation.

Solution

If $E(G) = \emptyset$, use x iff G can be recovered from x in c where x is either a word or the combination of several words

This gives two bidirectional notions of lexical interpretation and production.

Interpret w as G in c iff $G = \{f : p(f|c) \neq 0\}$ unless $E(G)$ is non empty and does not contain w.

To express G in c use a word w in $E(G)$ with maximal probability if $G = \{f : p(f|c) \neq 0\}$ or if $F(G)$ is empty, a word w (or combination of words) such that $G = \{f : p(f|c) \neq 0\}$.

The new proposal incorporates semantic blocking.

It also gives a solution to Grice's problem.

Mrs T. produced a series of sounds closely resembling the score of Home Sweet Home.

This is anomalous because it chooses a combination of words instead of a perfectly normal word with the same meaning: *w* and a high probability

Conclusion

A theory of (lexical) ambiguity

A. How to construct the MIU for a word that applies in the context?

add the context to p_w and check for blocking

mismatch: construct extension

B. What is the right word to use given that you want the hearer to understand it as a particular MIU

try to use $E(G)$ but check recoverability, else find a word or complex expression from which G can be recovered

C. How is the lexical knowledge structured to allow fast decisions on these two questions?

as p_w for interpretation
as E for routinised production

D. How can that structure be acquired?

by experience