COOKING WITH SEMANTICS

SCHIZO SCHOLARS

THE ACTS

ACT 1: THE PROBLEM

Understanding why instructions can be ambiguous and what those ambiguities are like by classifying the challenges faced.

ACT 2: THE SOLUTION

This act about the model proposed by the paper to deal with the challenges mentioned.

ACT 3: RESULTS

Some preliminary findings made by the authors of the paper "Cooking with Semantics"

ACT 4: APPLICATIONS

A list of applications as envisioned by the authors

ACT 1: THE PROBLEM

Expectations and outcomes

How ambiguities are caused by interpretation of situated text in instructions and why they matter for machine readability

Classifying the problems caused:

- Arguments
- Actions
- Control Structure

BUT...WHY?

Instructions unclear, ended up with the wrong yellow cake.

.... the uranium kind

We want to interpret instructions in a richer way to make them fit for automated systems.

Most instructions are easy to understand and follow along, but the instruction language being **situated** is a problem.

We wish to automatically interpret them into rich semantic representations.

Some of it is already marked in **DESSERT** -- **Document Encoding and Structuring Specification for Electronic Recipe Transfer**, but the instructions are still a block of text.

HUNGRY YET?

If not, well, now's a good time...

We'll use a Maggi Manchurian recipe as reference to support the points we make. As you'll soon see, most instruction-sets succumb to the problems we shall discuss.

So, let's go!

ARGUMENTS

You didn't tell me the ingredients were to be added *to the pan!*

- Elision
- The zero anaphora problem
- Implicit ingredients
- Incompletely specified arguments
- Unspecified preparatory steps

"अब टमाटर सॉस, चिली सॉस, सोया सॉस, नमक और काली मिर्च डालें. सॉस को 2-3 मिनट तक पकने दें."

"सबसे पहले, मैगी को वैसे ही पकाएं जैसे आप आमतौर पर बनाते हैं. पानी सूख जाने के बाद इसे दूसरे प्याले में निकाल लीजिए."

"एक दूसरे पैन में थोड़ा तेल गर्म करें और बचा हुआ अदरक-लहसुन, हरे प्याज़, कटे हुए प्याज़ और शिमला मिर्च डालें. इसे एक मिनट तक चलाते हुए भूनें."

"1/2 कप स्प्रिंग अनियन"

ACTIONS

Nobody told me I had to crack the egg before adding it!

- Actions have argument-dependent sense
- Sometimes they're implicit

"घोल को अपने पैन में डालें."

"कॉर्नफ्लोर डालें और हल्के हाथों से मिश्रण बना लें."

"2 पैकेट मैगी नूडल्स"

"1/2 कप स्प्रिंग अनियन"

"हरे प्याज़ से गार्निश करें"

CONTROL STRUCTURE

What do you mean "to taste"?!

- Info provided about sequence, coordination, alternatives and control conditions can be a bit complex.

"सबसे पहले, मैगी को वैसे ही पकाएं जैसे आप आमतौर पर बनाते हैं"

"सब कुछ अच्छी तरह से मिल जाने के बाद."

"इन बॉल्स को मध्यम आंच पर तलें. एक दूसरे पैन में थोड़ा तेल गर्म करें"

A COUPLE OF PREREQUISITES

Just so we're on the same page

Semantic role labeling:
Assigns **labels** to words or phrases in a sentence that indicates their **semantic role** in the sentence, such agent, goal, object, location, medium/instrument, owner, owned, etc

Case grammar:
Analyzes the surface syntactic structure of sentences by studying the combinations of semantic roles **(deep cases)**, which are required by a specific verb - for instance, "give" requires an agent, an object and a beneficiary.
Each verb selects a certain number of deep cases which form its **case frame**.

Frame semantics:
Extends case grammar. It relates linguistic semantics to encyclopedic knowledge. The basic idea is that one cannot understand the meaning of a single word without access to all the essential knowledge that relates to that word.

FUN FACT

Because why not :P

We have **one** person to thank for creating all three of the disciplines we just discussed (case grammar, semantic role labelling and frame semantics), and a lot more!

Charles J. Fillmore

ACT 2: THE SOLUTION

The method and the model

This act is about the actual model presented by the author and the processed involved in parsing recipes as specified.

Features:

- The machine will maintain a *latent* context that is dynamic but persistent.
- The model proposed has the overall structure of a *Markov Decision Process.*

MARKOV DECISION PROCESS

The model presented has the overall structure of a Markov Decision Process

This is a paradigm of decision making in situations where outcomes are partly random and partly under the control of the decision maker.

This makes sense because we're talking about it in the context of a machine doing things.

- We know what actions we want to take (the thing under control)
- The real world involves robots using actuators to move, so certain things are random in the sense that our actions may not exactly manifest themselves. That's where the randomness comes in.

MARKOV DECISION PROCESS

(continued)

The process follows the following basic idea:

- The process is at some state (stage) ***s***
- The decision-maker takes an action ***a***
- As a result, the process moves to state ***s'***

Important things to note:

- We're not assuming a deterministic world, i.e. the model only gives us the *probability* of getting to ***s'*** given the previous state ***s*** and the action ***a***
- The paradigm also tell us that ***s'*** is *exclusively* dependent on ***s*** and ***a***. and the states before absolutely do not matter.
- The process is Stochastic, i.e. it is wel described by a random probability distribution (at least assumed to be)

WHAT DOES THAT IMPLY TO COOKING?

Answering with an example of how this might work.

Suppose, we're thinking about a very simple recipe of creating saltwater from salt and water.

I have a glass of water
and a tablespoon of
sugar.
This is my state ***s***.

I pour the salt and stir.
(this is the action ***a***)

et voila! we have
saltwater.
This is state ***s'***

STATES AND ACTIONS

The notation and terminology used in the paper

State S (S_t):

The state, S_t is the current state of the world at time t. It represents the set of objects and attributes:

- ingredients
- containers
- quantity
- location
- condition (raw / cooked / minced etc.

Any object can undergo a change in state.

Action A (A_t)

As mentioned above, any object can undergo a change of state. An *action* causes said change and therefore can change the world state.

STATES AND ACTIONS

(continued)

We now present a model of the world

$$p(S_t \mid S_{t-1}, A_t)$$

This equation simply articulates the Markov decision model in the current context.
The state at t will be determined by the state at (t-1) and the action at t

To reiterate: it is a probability representation because we're not making an assumption for determinism.

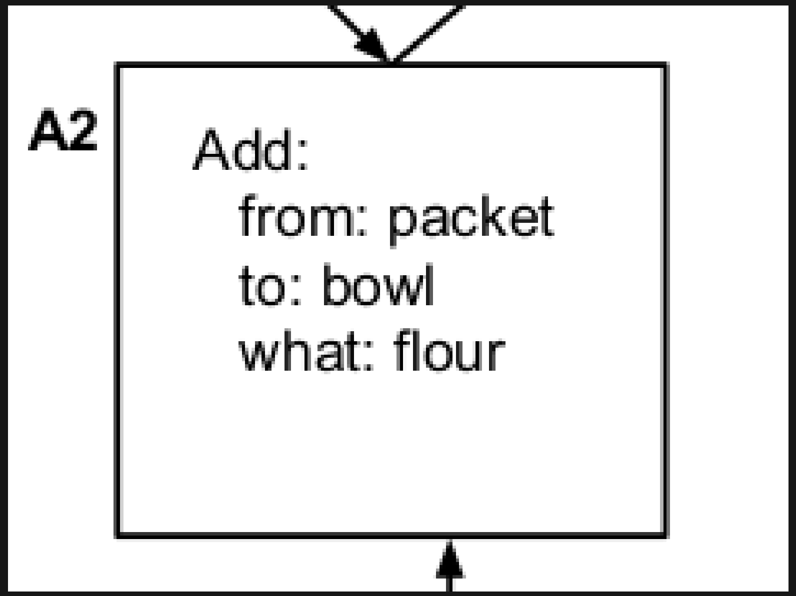
ENOUGH WORLD BUILDING...

Time to get real

That's it for how we'll represent the world to the machine.

Now, to parse the recipe text into this form.

SEMANTIC ROLE LABELING



Semantic role labels and getting semantic frames from the text will be essential.

Essentially, we take a verb and it's arguments together to make the semantic frame

In Hindi that generally involves using *karaka* relations.

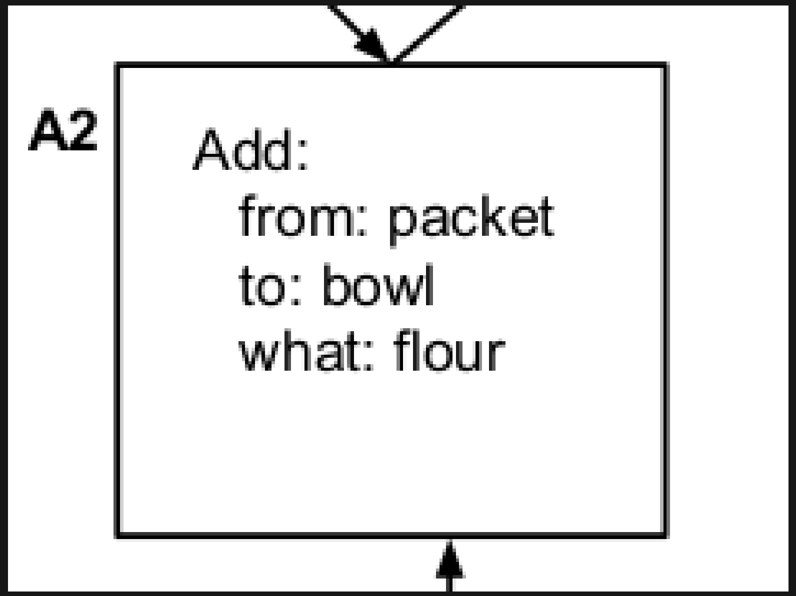
"dependency tree without the tree" - Abhinav S. Menon

"हल्के हाथों से मिश्रण बना लें"

बना:

मिश्रण (k2) k3 -से
हाथों (k3) (marker)

UNDERSTANDING ACTIONS



Action A (A_t)

Each action is represented by a *semantic frame.*

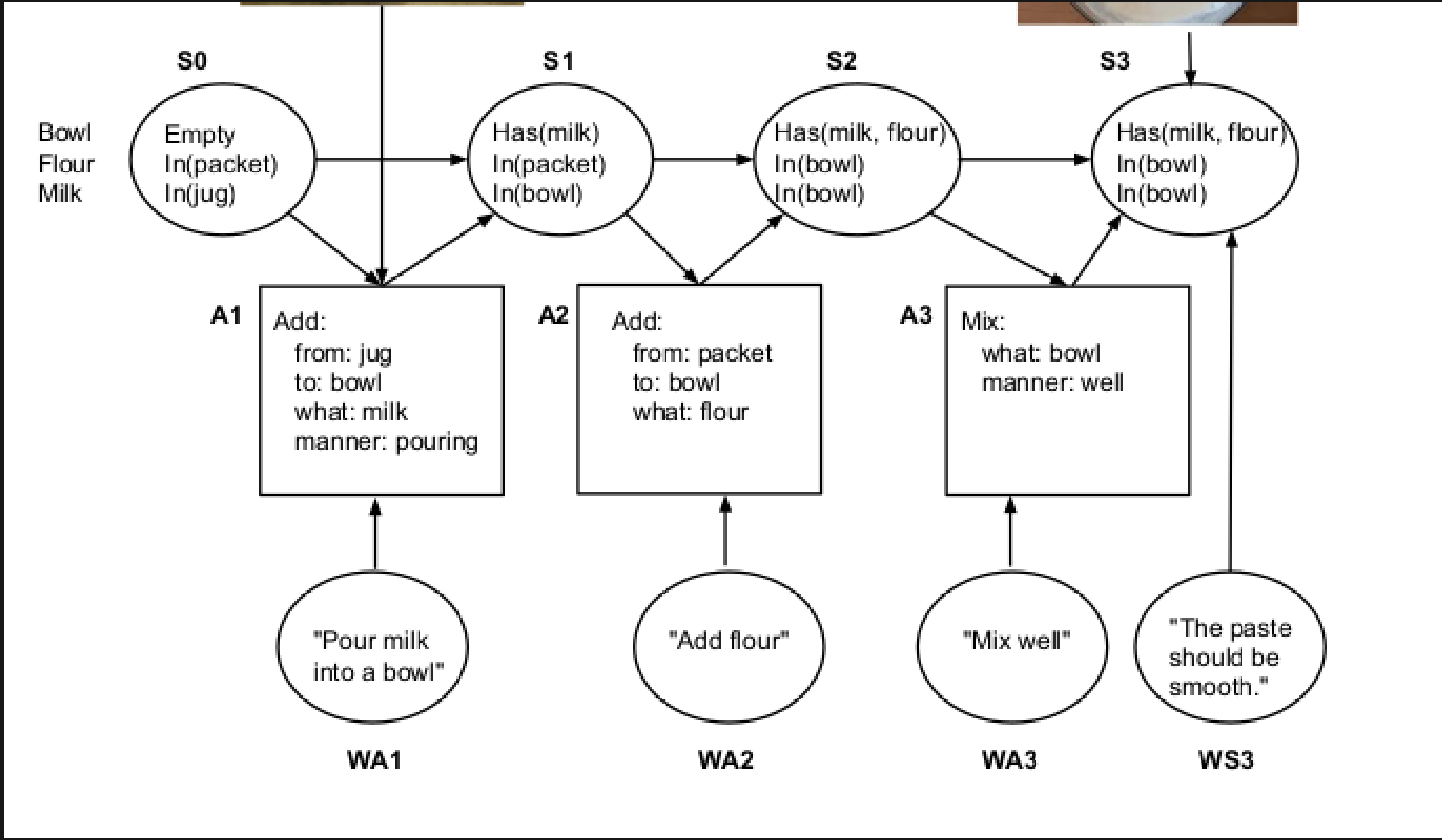
Essentially, we take a verb and it's arguments together to make the semantic frame

The paper assumes that the t^{th} sentence is the one referring to A_t and will be referred to as WA_t

Leading us to the conditional distribution

$$\mathbf{p(A_t \mid WA_t , S_{t-1})}$$

This basically tells us about $\mathbf{A_t}$ given the last state and the current sentence



S0
S1
S2
S3
Bowl
Flour
Milk
Empty
In(packet)
In(jug)
Has(milk)
In(packet)
In(bowl)
Has(milk, flour)
In(bowl)
In(bowl)
Has(milk, flour)
In(bowl)
In(bowl)
A1
Add:
from: jug
to: bowl
what: milk
manner: pouring
A2
Add:
from: packet
to: bowl
what: flour
A3
Mix:
what: bowl
manner: well
"Pour milk into a bowl"
"Add flour"
"Mix well"
"The paste should be smooth."
WA1
WA2
WA3
WS3

FrameNet search: "run"

Lexical Unit	Frame	LU Status	Lexical Entry Report	Annotation Report
run (through).v	Practice	Finished_Initial	LE	Anno
run away.v	Fleeing	Created	LE	
run off.v	Duplication	Created	LE	
run out.v	Expend_resource	Created	LE	Anno
run through.v	Cause_harm	Created	LE	
run-through.n	Practice	Finished_Initial	LE	Anno
run.v	Self_motion	Needs_SCs	LE	Anno
run.v	Leadership	FN1_Sent	LE	Anno
run.v	Impact	Created	LE	
run.v	Fluidic_motion	Finished_Initial	LE	Anno
run.v	Cause_impact	Finished_Initial	LE	Anno
run.v	Cause_motion	Created	LE	
run.v	Operating_a_system	New	LE	Anno
run.v	Path_shape	Created	LE	Anno
runway.n	Roadways	Finished_Initial	LE	Anno

UNDERSTANDING ACTIONS

(continued)

The frame-semantic parsing / semantic role labelling is done

Roles are filled in not just from the spans of the text, but also from the latent state vector.

The latent state vector is a representation of real world knowledge.

This knowledge is about the compatibility of candidate objects and the role labels of the frame of a word.
Like for instance, we can't whisk a solid egg

This is taken with information about which objects were recently mentioned.

We make guesses about the *zero anaphora* based on what was mentioned recently or acted upon recently

MAINTAINING STATE

In general we are deciphering Actions from the recipe and maintaining the state as it changes.

Often, recipes will describe the current state. In terms of what an object should look like afterr a given step.

We let this information be referred to by WS_t and use it to aid our prediction of S_t and use the function

$$p(S_t|WS_t)$$

EXTENDING TO VIDEO

The authors mention that often how-to videos are available in the recipe and that we would like to be able to use this information.

$$p(A_t | VA_t\ ,\ WA_t\ ,\ S_{t-1})$$

where VA_t represents the information retrieved from the video.

TRAINING DATA AND LEARNING

Most of the manipulations with all the probability functions mentioned before will be done by a machine learning algorithm.

It assumed that the training data is fully annotated with the A_t
The author expects to ease the particular assumption down the line.

It also assumed that S_0 i.e. the initial state is fully observed.

The world is assumed to be deterministic like a state machine. That is if we know the actions then theoretically expect to infer the sequence of states as they change exactly.

ACT 3: RESULTS

Expectations and outcomes

Some preliminary findings made by the authors of the paper "Cooking with Semantics"

HOW DOES IT FARE?

Comparing the proposed approach to the (then) current state of the art

The automatic tagger, which did not have access to the proposed latent context, often made errors in mapping verbs to the correct rolesets when compared to the manually annotated versions that contained explicit role labels as well as those created implicitly using the rich context.

Certain verbs lacked framesets entirely, like "prebake".

Others were segregated into incorrect rolesets.

52% of all verbs referenced at least one role from the latent dataset.

ACT 4: APPLICATIONS

Scope for usage in the real world.

DERIVING CANONICAL RECIPE

This model will allow us to take different forms of a recipe and unify them in some way into a "canonical form"

EXPLAINING INDIVIDUAL STEPS

The author proposes that at times a user can click a confusing step for clarification.
From what I understand of the paper, I don't think the model will be any good for elaboration to people.

ROBOTICS

I suppose this the most obvious application. Usage of Markov process makes this model very suitable for software that helps robots follow instructions written for humans.

REFERENCES

- https://www.mff.cuni.cz/veda/konference/wds/proc/pdf13/WDS13_105_i1_Brunetto.pdf
- https://storage.googleapis.com/pub-tools-public-publication-data/pdf/42245.pdf
- https://aclanthology.org/W14-3007.pdf
- Case grammar - Wikipedia
- Semantic role labeling - Wikipedia
- Frame semantics (linguistics) - Wikipedia
- food.ndtv.com/recipe-maggi-manchurian-hindi-955283