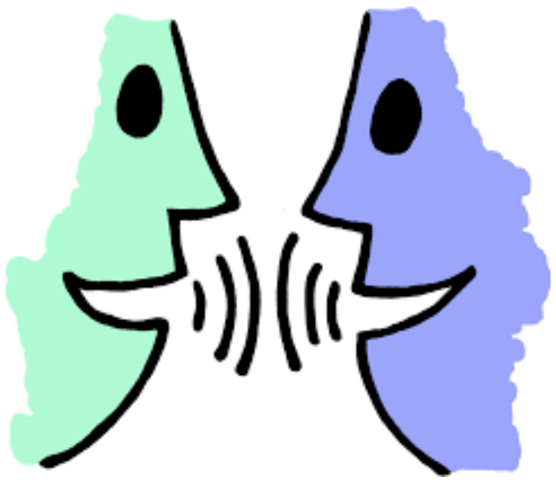
Cognitive NLP

The wondrous challenge of human language

Dave Raggett, ERCIM

11 January 2021

Why work on natural languages?



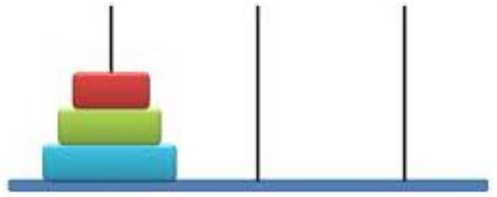
Courtesy of pixy.org

- Natural language is key to rich communication between humans and cognitive agents, just as it is now for communication between humans, and could be for communication between cognitive agents
  - *Natural language provides a shared means of expression in contrast to the internal representation of ideas within our brains which is unique to each of us in terms of concurrent patterns of activation of cortical nerve fibres*
- Natural language provides unlimited flexibility for expressing ideas, and can easily be extended with new words and phrases
- Natural language is also key to instructing cognitive agents as a solution to the manual knowledge engineering bottleneck
- Everyone should be able to use their preferred language for communicating with cognitive agents

Initial Experiments on NLP

- Use of text or speech to move discs in the towers of Hanoi game
  - https://www.w3.org/Data/demos/chunks/nlp/toh/
  - Initial proof of concept for shift-reduce parsing with chunks and use of rules to map syntax to semantics
- Dinner demo: dialogue between cognitive agents
  - https://www.w3.org/Data/demos/chunks/nlp/dinner/
  - Agents exchange chunks, whilst invoking speech API
  - Rules describe transitions between named tasks
- Parsing demo: tests that the shift-reduce parser is adequate for dinner dialogue utterances
  - https://www.w3.org/Data/demos/chunks/nlp/parsing/
- Ongoing work
  - Preliminary investigation on ideas for a new version of dinner demo where agents communicate using natural language
  - This will also feature use of causal reasoning for much greater flexibility and resilience compared to task transition rules

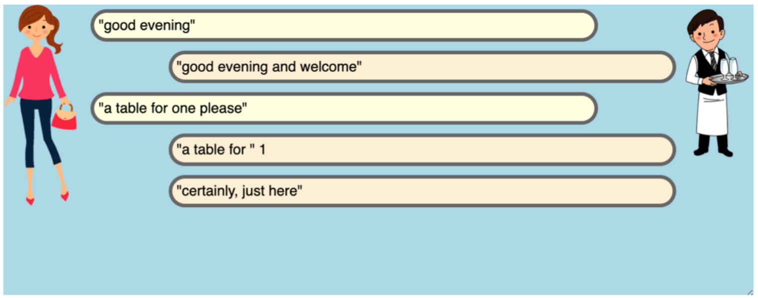
Towers of Hanoi



```
# move the red disc to the right peg
verb v1 {word move; subject p1; to p2}
phrase p1 {word disc; det the; adj red}
phrase p2 {word peg; det the; adj right}

# after application of ruleset
move m1 {disc disc3; to peg3}
```

Dinner Dialogue



Restaurant Knowledge Graph as chunks

```
# the names and sequence of courses
course {name starters; next main}
course {name main; next dessert}
course {name dessert}

# rudimentary dinner menu
menu-item {course starters; name prawn-cocktail}
menu-item {course starters; name tomato-soup}
menu-item {course starters; name mixed-salad}

menu-item {course main; name sea-bass; kind fish}
menu-item {course main; name plaice; kind fish; status unavailable}
menu-item {course main; name steak-pie; kind meat}
menu-item {course main; name spaghetti-bolognese; kind meat}
menu-item {course main; name vegetable-lasagne; kind vegetarian}

menu-item {course dessert; name apple-pie}
menu-item {course dessert; name ice-cream}

# drinks menu
drinks-menu-item {name sparkling-water; size bottle}
drinks-menu-item {name still-water; size bottle}
drinks-menu-item {name coca-coala; size bottle}
drinks-menu-item {name orange-juice; size bottle}
drinks-menu-item {name apple-juice; size bottle}
drinks-menu-item {name beer; size bottle}
drinks-menu-item {name red-wine; size glass, bottle}
drinks-menu-item {name white-wine; size glass, bottle}

# plan
dinner-plan {stage arrive, order, consume, pay, leave}

# wines
wine {colour red; grape cabernet-sauvignon}
wine {colour red; grape merlot}
wine {colour red; grape shiraz}
wine {colour red; grape pinot-noir}
wine {colour red; grape malbec}
wine {colour red; grape grenache}
wine {colour red; grape tempranillo}
wine {colour red; grape sangiovese}
wine {colour white; grape sauvignon-blanc}
wine {colour white; grape chardonnay}
wine {colour white; grape riesling}
wine {colour white; grape pinot-gris}

# water
sparkling-water kindof water
still-water kindof water
tap-water kindof water

# live status for tables and seats
table {location window; seats 4; status reserved}
table {location window; seats 4; status in-use}
table {location inside; seats 4; status in-use}
table {location window; seats 2; status available}
table {location window; seats 6; status in-use}
```

Ongoing activities

- Further development of chunk ontology for restaurant scenario to support causal reasoning
- Analysis for richer handling of syntactic details – but not really needed for dinner demo
  - e.g. inferring tense from verbs
  - Important for future demos that explore reasoning about time and for richer social interactions
- Analysis of dialogue acts for the dinner demo
  - Utterance: dialogue act + information
  - Relationship to semantic processing
- Exploring compositional approaches to syntax-semantic mappings for NLU/NLG
  - Implications of mirror system hypothesis*
- Survey of Syntactic-Semantic structures across spoken human languages
  - See Michael Campbell's analysis
- Unsustainability of manual development of mapping rules – need to understand how children develop language skills
- Macros as a baseline for mapping rules
  - Simple and surprisingly effective
- Statistical weights for selecting between competing rules
  - Influenced by situational ontology
- Compositional rules acting over syntactic components, e.g. nominal groups
  - Relationship to Fluid Construction Grammars
- Inducing rules from examples

* explained on next slide

Where did natural language come from?

The mirror system hypothesis is that human language was predated by development of the means for imitating complex actions by others, involving so called mirror neurons and Broca's area of the brain. Humans alone are able to recognize another's performance as a combination of more-or-less familiar actions and to use this recognition to approximate the action, with increasing practice yielding increasing skill.

- Humans appear to be unique in the infinite possibilities of our languages compared to other species
- Language evolved to help humans communicate about complex social activities, e.g. hunting
- Mimicking past hunts, communicating ideas for the next hunt, involving time and causation
- Imitation, body movements and dance with sounds as precursors to modern sign languages
- Overtime sound becomes dominant due to its convenience, with body language and facial gestures remaining as primeval vestiges
- Spoken language is over 300 thousand years old
- Written language just a few thousand years old
- Language as a social rather than logical medium

Photo used under Creative Commons from kevin dooley

Social Mimicry
it makes us human

Courtesy of snappygoat.com

- Mimicking others
  - Babies learning to smile from interacting with their mothers*
  - Children copying speech sounds of their peers (regional accents)
  - Learning how to say complex utterances by listening to others
  - Imitating dance movements of others on the dance floor or TV
  - Playing some music on a piano or guitar after listening to it
  - Choosing the same styles of clothes as your friends
- Socially driven
  - Emotionally satisfying, a feeling of belonging
- A common cognitive architecture
  - First, an internal model has to be learned from lower level sensory data, via increasing levels of abstraction, across multiple modalities
  - Second, you have learn how to map this internal model to a lower level model for motor control, via decreasing levels of abstraction, for execution by the cerebellum
  - Statistics for recognition of patterns is shared with their generation, e.g. shared across natural language understanding and generation
  - Incremental learning involving only weak supervision, and evolving effective models from many potential alternatives

* Smiling: there are 42 facial muscles and 6 muscles for moving each eye

Cognitive Science & Language

Emergence of compositional phonology, hierarchical phrase-structure and compositional semantics

Willem Zuidema, Institute for logic, language and computation, University of Amsterdam

Language is extremely complex, extremely useful, universally found in all human populations and fundamentally different from any system of communication in other animals. Language is also acquired spontaneously within a few years by every healthy human child, and, although there is a recognizable language processing network in the adult brain, it does not seem to rely on anatomical or neural structures, or cell types, proteins or genes that differ fundamentally from those found in closely related species without language.

These sets of properties of language pose a major challenge for the language sciences: how do we reconcile observations about the uniqueness of language with those about the biological continuity of the underlying neural and genetic mechanisms?

[Major Transitions in the Evolution of Language]

Prerequisites for Learning Languages

Adele Goldberg, Department of Psychology, Princeton

- Our pro-social motivation and skill (e.g., Hermann et al., 2007; Tomasello, 2008);
- the general trade off between economy of effort and maximization of expressive power (e.g., Levy, 2008; Futrell et al., 2015; Kirby et al., 2015; Kurumada and Jaeger, 2015);
- the power of statistical learning (Saffran et al., 1996; Gomez and Gerken, 2000; Saffran, 2003; Wonnacott et al., 2008; Kam and Newport, 2009);
- and the fact that frequently used patterns tend to become conventionalized and abbreviated (Heine, 1992; Dabrowska, 2004; Bybee et al., 1997; Verhagen, 2006; Traugott, 2008; Bybee, 2010; Hilpert, 2013; Traugott and Trousdale, 2013; Christiansen and Chater, 2016).

Frontiers in Psychology, 2016 – "Subtle Implicit Language Facts Emerge from the Functions of Constructions"

Cognitive NLP

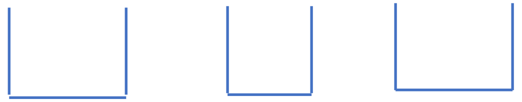
- Linguists focus on syntax rather than meaning, this is also true for recent work with Deep Learning, e.g. BERT & GPT-3, which use statistics over large corpora in lieu of meaning
- Cognitive AI supports the representation of meaning in terms of symbols and statistics
- Functional treatment of language in terms of a pipeline connecting different abstractions
  - *phonology, morphology, words, phrase structure, semantics* and *pragmatics*
- Understanding (NLU)moves along the pipeline from phonology to pragmatics
- Generation (NLG) moves along the pipeline from pragmatics to phonology
- A more realistic approach would include non-verbal communication as a side channel, e.g. facial gestures, eye, head, hand and arm movements
- However, written language shows that the latter aren't essential, although desirable for social interaction
- Cognitive constraints on working memory; incremental processing avoiding backtracking
- Evidence for separation of linguistic processing and general cognition
  - Observations of effects of dissociation given different kinds of brain damage
  - NLP is largely unconscious in nature and seemingly effortless – e.g. when deciding that a given utterance is ungrammatical in respect to languages learned fluently as a small child

ERCIM
European Research Consortium for Informatics and Mathematics

Limited Working Memory

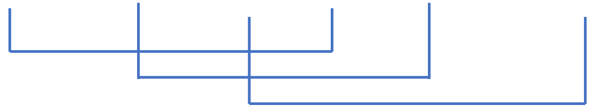
Same meaning, different languages:

EN: *that Ingrid saw Peter let Hans swim*



DE: *dass Ingrid Peter Hans schwimmen lassen sah*

NL: *dat Ingrid Peter Hans zag laten zwemmen*



Adapted from Christiansen and MacDonald (2009)

- Syntactic structure is related to word dependencies across an utterance
- Sequential, embedded and cross dependencies, see examples on left
  - Amenable to shift-reduce parsing using rules with a stack for working memory
- Human languages limit complexity, presumably due to constraints on working memory (e.g. stack/window size)
  - In the following, the first sentence is much harder to understand than the second

The apartment that the maid who the service had sent over was cleaning every week was well decorated.

The well decorated apartment was cleaned weekly by the maid sent over by the service.

Early and Late Binding

- Incremental parsing involves decisions on ambiguities
  - Expanding Abbreviations

    I'd love to see that vs *I'd seen that before*
    - Whether to expand 'd as "would" or "had" depending on the next word
  - Part of speech

    This book is great vs *Please send this to me*
    - Whether "*this*" is treated as a determiner or pronoun depends on next word
  - Word sense where the correct choice depends on the context
    - Priming effects modelled using spreading activation
  - Anaphoric and cataphoric reference, e.g. pronouns
    - Jane entered the room. She was wearing a blue dress.
    - When he arrived, John noticed that the door was open.
    - She was reading her new book. It was a novel by Charles Dickens.
    - When Peter got to the door, he found it was unlocked.
  - Prepositional attachment
    - Jane was given a necklace by John for her birthday.
- Sometimes a decision can be made based upon the preceding words
- Sometimes you have to defer a decision until further words have been processed
- Avoid backtracking
  - Synchronous local resolution by peeking at next word
  - Non-local resolution involving asynchronous feedback from layer above

Episodic Contexts

The meaning of utterances is recorded in chunk contexts* as a dialogue history distinct from general semantic knowledge.

This supports search across previous sentences when interpreting pronouns and nominal groups, and forms part of episodic memory.

The best binding may depend on gender, semantic match, adjacency and statistics, but should be obvious to both the person speaking and the person listening.

* Analogous to RDF named graphs

End-to-End Communication Testbed

- Plan for work on a proof of concept web-based demonstrator for end-to-end communication via natural language with a focus on meaning
- First select an utterance from a suite of utterances
- Each utterance is expressed as a simple chunk graph that describes its meaning in terms of semantics and pragmatics
- This is stochastically transformed into a chunk graph for the corresponding syntactic phrase structure, and thence into a sequence of words
- The words are then mapped back to the phrase structure and meaning using concurrent processing of syntax and semantics without backtracking
- The meaning is then compared to the original utterance
- This will demonstrate round-tripping of meaning through NLG and NLU

Example

- Meaning:
  - give act1 {from p1; to p2; thing t1; time past}
  - person p1 {gender male; name "John"}
  - person p2 {gender female; name "Mary"}
  - book t1 {quantity 1}
- Phrase structure:
  - vp _:1 {verb give; tense past; subject _:2; object _:3; to _:4}
  - np _:2 {noun john}
  - np _:3 {det a; noun book}
  - np _:4 {noun mary}
- Resulting text utterance:
  - John gave a book to Mary
- Alternatives involving different emphasis
  - John gave Mary a book
  - Mary was given a book by John
  - A book was given to Mary by John

- The meaning includes the speech act – in this example – an assertion as the default speech act
- The model is easily extended, e.g. to indicate the primary and secondary emphasis
- The meaning is mapped to a phrase structure via a sequence of paired rules reflecting compositional semantics for natural language
- These rules are stochastic reflecting the statistics observed in previous dialogues
- Rules have conditions and actions that enable cascaded execution across rule sets
- Rules are paired across meaning level and phrase structure level with shared variables
- Rules are reversible to support NLU and NLG
- Future work to examine use of stock phrases and metaphors as part of everyday language

Natural Language Generation

- Start with a chunk graph expressing communication intent
- Progressively map to syntactic structure and word sequence
- Treatment of pronouns and nominal groups* avoiding ambiguity

  Good: *I love my dog. I take it on long walks.*

  Bad: *Wendy likes Janet. She has black hair.*
- Use Gricean principles for communication between collaborative agents, e.g. avoiding repetition and taking into account shared knowledge – this involves a record of dialogue history
  - minimise speaking effort whilst avoiding misunderstandings
- Merge separate statements into a compound statement if appropriate

* nominal groups refer to models of things introduced in the dialogue, e.g. "a red door"

NLG for the Example

- Suite of stochastic mapping rules covering
  - regular verbs, dealing with the subject and object
  - tense and number agreement
  - nominal group for a named person
  - nominal group for a thing with a quantity
  - prepositional phrases (e.g. *to* and *by*)
- These rules cascade to complete the mapping
  - The application of a rule triggers other rules
  - Overall structure first, fill in details later
- Pronouns as references to nominal groups
  - Anaphoric references: if the nominal group has already been used earlier in the same utterance, or recently in the dialogue, provided the pronoun would be sufficiently unambiguous
  - Cataphoric references: Only sensible if there are just two occurrences, use pronoun for 1st occurrence and nominal group for 2nd

- Meaning:

  give act1 {from p1; to p2; thing t1; time past}
  person p1 {gender male; name "John"}
  person p2 {gender female; name "Mary"}
  book t1 {quantity 1}
- Phrase structure:

  vp _:1 {verb give; tense past; subject _:2; object _:3; to _:4}
  np _:2 {noun john}
  np _:3 {det a; noun book}
  np _:4 {noun mary}
- Resulting text utterance:

  John gave a book to Mary
- Alternatives involving different emphasis

  John gave Mary a book
  Mary was given a book by John
  A book was given to Mary by John

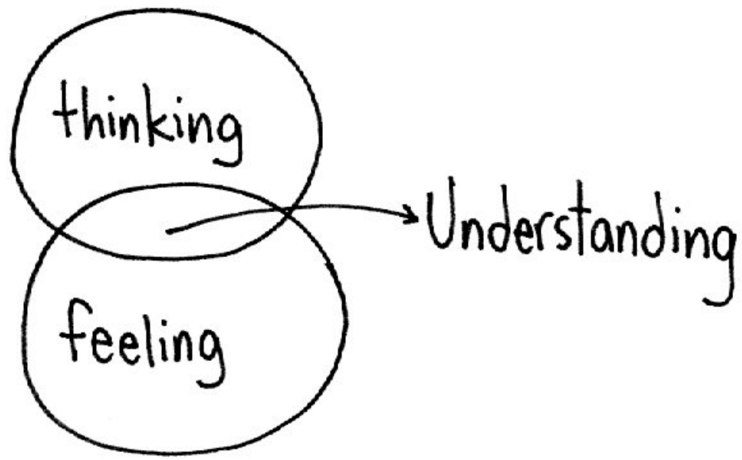
Lots of work needed on the details, including an easy to understand rule syntax, parser and engine

Natural Language Understanding

Concurrent incremental processing of syntax and semantics

1. Normalisation of text stream
   I'd like tomato soup ⇨ [i, would, like, tomato, soup]
2. Part of Speech & Word Sense
   Use of lexicon and spreading activation: tomato ⇨ noun
3. Syntactic Phrase Structure
   vp _:1 {verb would, like; subject _:2; object _:3}
   np _:2 {pron i}
   np _:3 {noun tomato, soup}
4. Syntactic Details
   Analysis of tense, mood, number, person, …
5. Syntax to Semantics
   Use of rules to map to semantic chunk graph
6. Dialogue context
   Use of dialogue context to infer intent
7. Goal driven reasoning
   e.g. decide which dish to order for main course

Understanding proceeds at multiple levels concurrently as a pipelined process that minimises demands on working memory



Understanding at a social level involves feelings as well as thinking – this is something to aim for in future with planned integration of functional models of the limbic system

Part of Speech for Word Classes

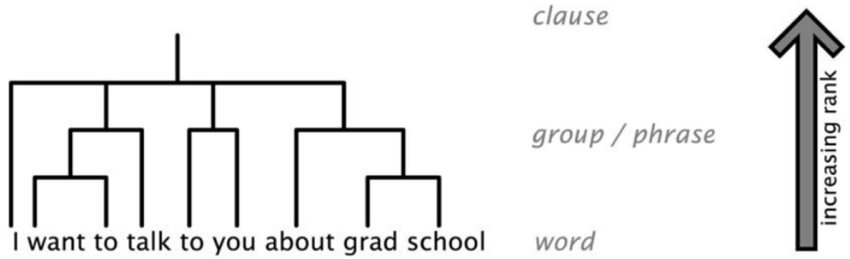
- Native speakers can instantly recognise that an utterance is ungrammatical, how?
- Statistical models can explain regularities based upon word classes, e.g. for English
  - noun, verb, adjective, pronoun, preposition, conjunction, interjection and determiner
- Some words are in multiple classes, but each occurrence in an utterance involves a particular choice of class
- In principle, word classes could be induced from a large corpus of utterances given suitable criteria
  - Note that it is much harder to get extensive corpora of ungrammatical utterances, so learning algorithm* seeks to "explain" all grammatical utterances in corpora
- However, there are trade-offs to consider

  A large set of classes may increase accuracy in predicting grammaticality, but at the cost of increased complexity for parsing, and difficulties in identifying the correct word class in any given instance.

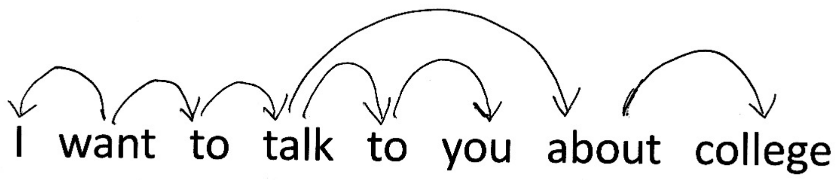
- Working hypothesis is that a small set of parts of speech is sufficient, and can be supplemented by additional lexical properties where needed
- Natural language generation can be guided by statistics from natural language understanding, reducing the likelihood of generating ungrammatical utterances
  - Humans like to mimic other humans
- Part of speech tagging by computer often uses between 50 to 150 classes for English, however these are difficult for non-experts
- It seems better to stick to the part of speech classes used in everyday English dictionaries
- A few extensions such as numbers, dates and times, and special treatment for common ways to express units of measure, e.g. *6m* for 6 metres
- Option to use supplementary lexical properties, e.g. to distinguish cardinal vs ordinal numbers

* See work on **learnability theory** on the challenges posed by lack of negative examples

Phrase Structure



From [Kappagoda, 2009]



I want to talk to you about college
vp _:1 {verb want; subject :_2; to :_3}
np _:2 {pron i}
np _:3 {noun talk; to _:4; about _:5}
np _:4 {pron you}
np _:5 {noun college}

- Syntactic word classes (part of speech) allow utterances to be segmented
- These segments can be arranged in a tree
- The attachment point depends on both syntactic and semantic considerations, which therefore need to be processed concurrently
  - Person wanting some thing or action
  - Talking as an action
  - Talking to someone
  - Talking about some thing or action
- Semantic agreement overrides adjacency for the attachment of prepositional phrases
- Semantic mapping rules are designed or learned for the structures generated by the phrasal analysis
- Phrase structure as data stream for understanding meaning in terms of semantics and pragmatics

Phrase Parsing as a stage in NLU pipeline

ERCIM
European Research Consortium for Informatics and Mathematics

- Earlier stages expand abbreviations, strip punctuation and coerce to lower case, as a means to emulate listening to spoken text
- Coarse syntactic groupings and relationships, deferring details until following stage
- Chunks for verb phrases, noun phrases, prepositions, conjugations, etc.
- Each chunk as an n-ary term
  - chunk types are informal
- Lists for multiple instances
  - e.g. adjectives, nouns, verbs
- Prepositions are reduced using their name as a property, see facing example with *for*
- Devolved disambiguation
- Readily extended as needed
- Future work could reformulate parsing in terms of shift/reduce rule language

```
# are you ready to order

vp _:10 {verb are; object _:12; subject _:11;
intent question; to _:13}
np _:12 {adj ready}
np _:11 {pron you}
vp _:13 {verb order}

# what would you like for the main course

vp _:15 {verb would, like; subject _:14;
object _:16; for _:18}
np _:14 {pron what}
np _:16 {pron you}
np _:18 {det the; adj main; noun course}
```

Macros for mapping syntax and semantics*

As a baseline for more sophisticated approaches

- There can be many ways to state the same thing, e.g. to order a dish named “steak pie”
  - The steak pie
  - I would like the steak pie
  - I will have the steak pie
- This can be described with macros with embedded variables e.g.
  - the <main-dish>
  - I would like the <main-dish>
  - I will have the <main-dish>
- Macros could be grouped by the context and dialogue intent, together with statistical information to help when choosing between competing matching macros
- Macro engines can improve robustness, using chunking to skip over unimportant text
- Macros can be combined with rules to convert variables to different forms, e.g. from “one” to the integer 1
- Macros can have multiple variables, e.g.
  - a <measure> of <drink>
- Macros could be learned from examples, e.g. by matching words like “glass” and “cabernet sauvignon” against terms in an ontology, and generalised conditions, e.g. *wine:grape*
- But there comes a point where macros won’t scale well and a more powerful rule language is needed
  - Enabling complex conditions for variables
  - Exploiting compositionality of language
- We thus need to look beyond macros

* Macros are used for W3C’s zakim and rrsagent IRC bots for use in managing teleconferences

Related Work

- Systemic Functional Grammar
  - Originated by Michael Halliday
  - Linguistics as part of sociology
- Grammatical systems are seen as closed, i.e. having a finite set of options
- Lexical sets are open with new words appearing regularly
- Language as personal exercise of choices in expressing meaning
  - Human experience – ideation
  - Interpersonal – dialogue actors, social distance and relative social status
  - Text – organisation and communicative nature

- Fluid Construction Grammar
  - FCG does not make any claims about biological realism or cognitive relevance
- Rules relating utterances and meaning for use in NLU and NLG
  - Rules referred to as construction schemas
  - Conditions with variables and unification
  - Actions create or update models
- Contrast with generative grammars (Chomsky) where syntax, semantics and pragmatics are each self-contained
  - FCG rules span syntax and semantics
- FCG focuses on language usage by individual speakers rather than abstract notion of grammatical competence

How do these approaches address disambiguation?
Does it involve unification, inference, statistics, other?

Universal Grammar

- Noam Chomsky proposed in the ‘50s and ‘60s that humans have an innate capability for learning language that is based upon a pre-wired ***universal grammar***
  - Justified by putative claims as to the *poverty of the stimulus*, e.g. near absence of negative examples, making it difficult to learn complex grammar rules
- Each sentence in a language has two levels of representation: a deep structure and a surface structure
  - Chomsky argued for the mathematical treatment of grammar independent of meaning, essentially ignoring the role of language for social communication
  - Formal theory of language across a population of speakers as a whole that determines whether a give utterance is grammatical or ungrammatical
  - Transformation Grammar: application of phrase structure rules followed by transformation into the corresponding surface structure
- Unfortunately, all attempts to identify this universal grammar have failed when confronted with the variations across human languages
  - This includes sign languages as well as spoken languages
- Most theorists now believe that humans have innate cognitive abilities that enable infants to progressively learn language as layers of abstraction
  - Statistical learning with weak supervision as we build hierarchical models to explain what we hear in the context of social interaction and shared attention
  - Building upon our ability to handle temporal sequences

Noam Chomsky claimed that all languages contain similar structures and rules, noting that across the world children acquire language seemingly effortlessly, suggesting that we're born with the basics already present in our brains.

Operational vs Formal Semantics

- It is commonplace to approach the meaning of language by adopting first order logic as a formalism
- That sounds reasonable at first glance – a reductionist approach has served science well – and logic has a long philosophical tradition dating back to classical Greece
- But it doesn't work as an adequate model of human reasoning

  "Most people like ice cream" isn't first order as it involves counting, statistics and fuzzy context dependent concepts
- Philip Johnson-Laird studied how people deal with real or imaginary situations
- He showed that people don't use a logic made up of formal or explicit rules of inference
- We instead construct mental models based upon what we know, e.g. working with existing or imagined examples
- Moreover, human reasoning is related to emotions, something that both Minsky and Johnson-Laird have noted
- Operational semantics involves a blend of graph data, statistics, rules and graph algorithms
- Reasoning occurs at multiple levels based upon what has worked for us in past experience
- It can lead to faulty inferences, but by and large works well enough for everyday needs
- You learn from your mistakes

Models of Disambiguation

- To avoid backtracking, we need to identify the best word sense and part of speech based upon the preceding words and the immediate next word
- Likewise, we need to identify the best attachment point after parsing a prepositional phrase
- This process of disambiguation involves graph algorithms that take the semantic context into account
- This is usually an unconscious process, and as such shouldn't burden the cortico-basal ganglia rule engine

- Cognitive databases mimic human memory with spreading activation, exponential decay and stochastic recall
- This can account for priming effects via spreading activation across connections between the lexicon, dialogue history, episodic and declarative knowledge
- The parser can propose candidate attachment points for prepositions
- These candidates are asynchronously ranked and the best one wins

Further progress depends on identifying suitable demo scenarios

Dialogue Acts

- Small, but open set of dialogue acts
  - e.g. questions, answers, explanations, commands, assertions, exclamations
- General assumption that people are normally acting cooperatively in social situations
  - Grice's maxims
- Yes-no question usually answered with accompanying information
  - As appropriate to the inferred reason for asking this particular question at this time
  - Or as explanation of why the answer was in the negative
- Recognising dialogue acts from utterances
  - Explicit: e.g. with a question such as *"could I sit by the window"*
  - Implicit: e.g. when an assertion is given in response to a yes-no question
  - Via an assertion + signal such as *"a table for one **please**"*

c: good evening # greeting
w: good evening and welcome # acknowledgement
c: a table for one, please # yes-no question
w: certainly # acknowledgement
W: just here # deictic reference as response to question
c: could I sit by the window # yes-no question
w: I am sorry # negative acknowledgement
w: the window tables are all reserved # explanation
c: this table will be fine # acceptance of earlier suggestion
w: are you ready to order # yes-no question
c: yes # acknowledgement
c: I will have the tomato soup for starters # assertion
…

Semantic Interpretation

- For some previous work see:
  - *From Syntax to meaning in Natural Language Processing*, Alexander Hauptmann, 1991
  - *Syntax-Semantic Mapping for General Intelligence*, Lian et al, 2012 (for work on OpenCOG in terms of hypergraph homomorphism)
- Hauptman describes rules using partial syntactic structures for conditions, where the actions specify add/delete operations on semantic graphs expressed as frames
- Frames are similar to chunks in the sense of being collections of properties that can reference other frames
- The semantic interpretation may involve additional reasoning about the dialogue context and situational plan using the semantic mapping as an input
- Are the mapping rules implemented by the Cortico-Basal Ganglia circuit?

```
# consider “this table will be fine”
vp _:28 {verb will, be; subject _:27; object _:30}
  np _:27 {noun table; det this}
  np _:30 {adj fine}

# the chunk parse tree could be mapped to
# an enhanced syntactic representation
be {@subject r1; @object fine; time future}
reference r1 {thing table; location near}

# a possible semantic representation given
# the task of finding a table is as follows.
# It recognises that this task is now done
# so we can move on to the next in the plan
accept {thing table; location near; @leave find-
table; @enter review-menu}
```

Learning rules from examples

- Hauptmann notes that computer code for transforming surface representations into deeper meaning representations is hard to write and debug – you need to be familiar with both the grammatical mechanisms and the domain models
- Rule languages can simplify this task, but are still dependent on human experts
- We ultimately want cognitive agents to learn the rules for themselves by generalising from what they have heard in conversation
- We need to find modest stepping stones towards that goal
- Hauptmann's work is an inspiration, and I am continuing to look for other such work, and a better understanding of the Fluid Construction Grammar framework

- One idea is to use a collection of examples with complete or partial utterances, and their semantic representations
- The agent uses what it already knows to parse the example utterances, thereby automating adaption to changes in the parser
- This assumes the lexicon and an ontology which can be used to search for matches and as a basis for generalisations
- Utterances are used to search for existing rules which could be generalised, specialised or combined to handle them
- Opportunity for applying reinforcement learning
- I am now working on compiling concrete examples to drive further analysis

I am also talking with [Lise Fontaine] at University of Cardiff's Centre for Language and Communication Research, where her focus is on meaning using Michael Halliday's work on Systemic Functional Grammar

Training Examples

These are used in conjunction with a suitable lexicon and knowledge graph

I’ll have tomato soup for starters
order {course starters; choice tomato-soup}

a glass of red wine
order {drink red-wine; measure glass}

Would you like anything to drink?
wh-question {order drink}

Are you ready to order?
yn-question {task order; status ready}

Oh dear
exclamation {conveys sadness}

This mapping assumes an ontology that defines available choices for each course. Likewise for the drinks menu and measures. These examples generalise to other dishes, courses, drinks and measures. *When does such generalisation take place, i.e. eagerly or lazily?*

yes-no questions can be interpreted as such or as a request for information, i.e. a wh-question. The ontology could distinguish ordering food and drink, and “to eat” as denoting food.

The ontology could enumerate a range of different kinds of feelings, e.g. for dishes.

How could we use training examples to learn about attachments for prepositions? This relates to incremental semantic processing of phrases and nominal groups. It also points to opportunities for acquisition of factual knowledge during dialogues.

Machine Learning

- Semantic interpretation is compositional, i.e. later rules build upon work by earlier rules
- Training examples should reflect how parser invokes semantic processing at different stages of parsing a complete utterance
- Desirable to be able to serialise rules for initialisation and debugging machine learning, as well as initial exploration of ideas
- Mapping rules have similar requirements to chunk rules (cortico-basal ganglia engine)
  - Rule conditions act on chunk trees
  - Rule actions create and modify chunks
  - Conditions and actions may invoke graph algorithms, e.g. to test against the ontology or dialogue history (if provided)

I # signalling the person speaking the utterance
I'll have # signalling a choice
tomato soup # signalling a dish from the dinner menu
for starters # signalling a course
I'll have tomato soup for starters # putting it all together

- Heuristics for rule generalisation, e.g.
  - John loves Mary ⟶ <person1> loves <person2>
- Rule indexing for searching for existing rules that could be combined or adapted via toolkit of heuristics
- Reinforcement learning for rule chains
  - Success/failure at semantic processing an utterance is used to update scores for all rules used to process an utterance
- Next step is to work through some examples

Rules and Reasoning

Following Hauptmann and work on FCG*

- Semantic interpretation is linked from the corresponding syntactic structures
  - e.g. from noun phrases and verb phrases
- **Conjunction of conditions** as partial syntactic structures
  - Chunks constraining what they match
  - Named variables with optional types
  - Unification as per construction grammars
- **Sequence of actions** that specify update, add and delete operations on semantic chunk graphs
- Statistical info to determine importance of rules in this context
  - Used to resolve conflicts
  - Shared across NLU and NLG

- Rules can invoke **named algorithms**, e.g.
  - Applying type constraints for variables
  - Scoring semantic compatibility
  - Mapping pronouns to individuals or groups
  - Mapping nouns to things in this episode
  - Scoring potential attachment points for prepositional phrases
- Challenges for use in both NLU and NLG
  - Declarative relationships that can be applied directly for NLU and NLG
  - Paired functions and inverse functions
  - Unpaired functions that can only be used one way for NLU only or for NLG only
- Ongoing study to better understand the different kinds of functions needed for the restaurant scenario, and how they could be generalized for other scenarios

* Fluid Construction Grammars

Further Ideas

- Training
  - Creates rule sets describing how semantics models are generated given current state of parsing and associated syntactic and semantic models
  - Rules are proposed/updated using a collection of heuristics
- Run-time
  - Stochastically select matching rules
  - If rule succeeds boost its strength and likewise for the heuristics it is based upon
  - Propagate score to previous rules as part of reinforcement learning
  - Update NLU statistics for use in NLG
- Adaptation
  - Meta-rules for generalising existing rules when an utterance is found similar to previous ones

- Linguistic rules annotate syntactic model with semantics, i.e. link from syntax to semantics
- Rule conditions describe tests which may involve the lexicon, episodic context and ontology
  - Direct matching + variable bindings
  - Inference, e.g. testing class membership, testing episodic facts
  - Statistics – e.g. priming effects
- Rule actions describe how to create or update associated semantic models
- Possibility of invoking cortico-basal ganglia rule engine for conscious processing
  - Sequential rule execution against module buffers
  - Goal and task driven, see robot and smart home demos

Summary

- Cognitive AI seeks to mimic human memory and reasoning at a functional level
  - Contrast with logic and formal semantics
- Chunks are more convenient than RDF and add sub-symbolic information on past usage
  - Semantic Pointers in n-dimensional spaces
- Chunk rules for cortico-basal ganglia circuit where conditions and actions apply to cortical buffers; actions can also invoke cortical algorithms and initiate actions that are followed through by the Cerebellum
- Future plans for modelling Limbic system
  - Integration of emotions and cognition: understanding through thinking + feelings
- Open question as to how NLP works alongside the sequential rule engine embodied by the cortico-basal ganglia circuit
  - What kinds of linguistic processing can be handled unconsciously and executed concurrently with conscious thought?
- Concurrent syntactic and semantic processing to address ambiguities and avoid backtracking
- Shift-reduce phrase structure parsing followed by further syntactic analysis
  - Small set of syntactic word classes
- Chunk contexts for episodic memory and dialogue history analogous to named graphs
  - Plus tasks for reasoning about plans
- Separate rule language for syntax-semantic mappings and pragmatics
  - Conditions over phrase structures
  - Actions that create/update semantic structures
  - Functions that invoke graph algorithms
- Machine learning of mapping rules from examples and generalizations thereof
  - Compositional approach to semantics
- Shared statistics between NLU and NLG
  - Implications of the fan effect

Where next?

Cognitive Natural Language Processing

social communication and a solution to the manual programming bottleneck

- Cognitively plausible processing model
- Incremental word by word concurrent syntactic and semantic processing without the need for any backtracking
  - Reducing load on limited working memory
- Use of statistical information to guide choices, e.g. for priming effect on word senses
  - Disambiguation on behalf of parser
- Offloading cognition with graph algorithms
  - Freeing up the cortico-basal ganglia circuit, and escaping the limits it imposes
- Separation of concerns
  - Word sense and part of speech
  - Referents for nominal groups*
  - Attachments for prepositions
- Simple robust shift-reduce parsing with a loose grammar and limited parts of speech
  - Parse tree and lexicon expressed with chunks

* a group of words which represents or describes an entity, and generally synonymous to a noun phrase

- Followed by processing further syntactic details such as time, person, mood, possessive, determinate, number, gender, …
  - Modelled as chunk properties
- Meaning in context of dialogue goals and task status, e.g. typical steps for visit to restaurant
  - Situational ontology + lexicon
  - Compositional semantics
- Declarative rules for bidirectional mapping between syntactic and semantic descriptions
  - Dialogue acts
  - Associated information
- Operational semantics, where the meaning relates to how cognitive agent behaves in a given context
  - Avoiding limitations of first order logic
- To gather statistics from NLU (competence) for use in NLG (performance)
- To learn from experience, generalising from examples