Fast and Frugal Detection of Chiastic Protofigures in English Subsection of CHILDES Corpus

regex strikes back

Daniel Devatman Hromada[123]
daniel@wizzion.com

[1]Université Paris 8 / Lumières
École Doctorale Cognition, Langage, Interaction
Laboratoire Cognition Humaine et Artificielle

[2]Slovak University of Technology
Faculty of Electronic Engineering and Informatics
Department of Robotics and Cybernetics

[3]Universität der Künste
Fakultät der Gestaltung, Berlin

Table of Contents

Computational (Developmental) Psycholinguistics

C(D)P

Is a cross-over between computational linguistic (and/or Natural Language Processing) and developmental psycholinguistics.

Main objectives:

1. use computational methods (data-mining, information retrieval, NLP etc.) to gain novel insights about ontogeny of language competence in human children
2. develop computational models of language acquisition and embed them into language-interacting artificial agents

In this talk we focus solely on the first objective.

CHILDES

CHILDES corpus: a gem of gems

Child Language Data Exchange System (MacWhinney&Snow, 1985)
http://childes.psy.cmu.edu/data
http://wizzion.com/CHILDES/ (mirror from 6th Feb 2016)

1. more than 50 years of tradition
2. more than 1.5 GigaBytes of mostly textual data contained in cca 30000 transcripts
3. at least 26 languages, dialects or language combinations
4. Creative Commons BY-NC-SA licence

CHAT format

CHAT system provides a standardized format for producing computerized transcripts of face-to-face conversational interactions. (MacWhinney, 2016; http://childes.talkbank.org/manuals/chat.pdf).

```
@Languages:     eng
@Participants:  CHI Eve Target_Child , MOT Sue Mother , FAT David Fathe
@ID:    eng|Brown|CHI|1;6.|female|||Target_Child|||
@ID:    eng|Brown|MOT|||||Mother|||
@ID:    eng|Brown|COL|||||Investigator|||
@Date:  29-OCT-1962
*MOT:   one two three four .
%mor:   det:num|one det:num|two det:num|three det:num|four .
%act:   tests tape recorder
*CHI:   one two three . [+ IMIT]
```

A non-negligeable advantage

Majority of transcripts follow the principle: ONE LINE = ONE UTTERANCE.

Computational (& Cognitive) Rhetorics

Computational Rhetorics

A discipline which has attained its maturity at Computational Rhetorics Workshop organized by Harris and Di Marco at University of Waterloo.

Computational-Cognitive Rhetorics

A disciplne using computers to better understand why rhetorics casts such a powerful curse on human minds.

Computational-Developmental Rhetorics

Using computers to elucidate the process of **ontogeny of rhetoric competence in human children**.

"Child's spontaneous remark is more valuable than all questioning in the world." (Jean Piaget)

Main concept(s)

Scheme

A scheme is a generic form which corresponds to one or more distinct constellations of observables.

Regular expression

A sequence of characters that defines a search pattern.

Perl-Compatible Regular Expressions

Concise and expressive regex standard. Much more powerful than regular grammars: it is possible to perform back-tracking!

Backtracking

Allows us to match that, which has already been matched: paves the way to detection of repetitions.

Main idea(s)

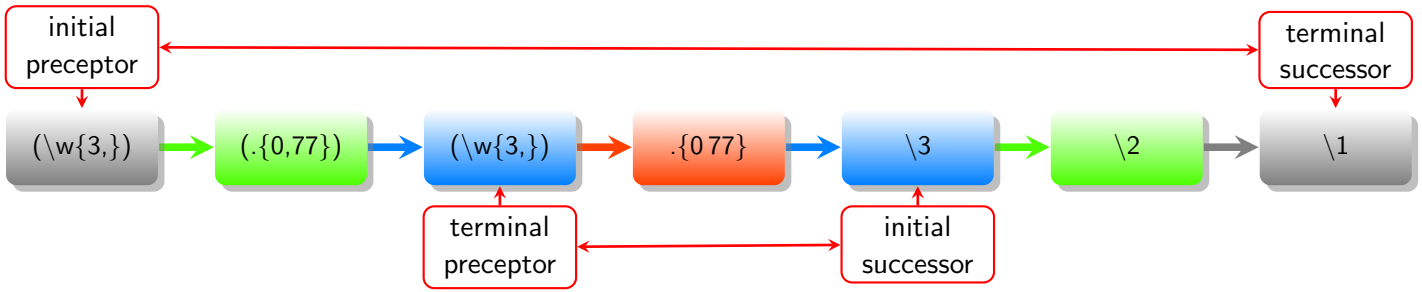
Main idea

Chiasms are repetition-based schemata $A_1 B_1 C_1 X C_2 B_2 A_2$ (or $A_1 B_1 X B_2 A_2$).

Note that the presence of middle term (B) and separator term (X) can be considered as facultative. But in order to detect chiasm, the initial preceptor (A_1) has to be strongly reminiscent (and ideally identic) to terminal successor (A_2). Idem for relation between terminal preceptor (C_2) and initial successor (C_1).

Table of Contents

Regex implementing the main idea



Note that nodes of a chiasmatic structure form a double-closed graph.

Demo

Run this shell command* :

```
grep -irP '^\*MOT:.*(\w{3,}) (.{0,77}) (?!\1)(\w{3,}).{0,77}\3 \2 \1' *Eng*
```

in the directory into which You downloaded and unpacked the CHILDES corpus.

Note that the extractor can be parametrized with change of numeric values: e.g. changing (\w{3,}) to (\w{1,}) could potentially allow You to detect grapheme-level metatheses like "*asteriks with an asterisk*".

* Regex sequence is hereby transfered to Public Domain under Creative Commons BY-NC-SA (Author Attribution, Non-Commercial, Share-Alike) licence.

You'll see many playful ones...

- pear pear yummy yummy yummy yummy pear .
- my name is Joey Joey Joe Joe Joe Joe Joey .
- I think I can I think I can I think I can I think I can I think I can I think I can I think I can .
- tick tick tick tick tick tick tick tick tick tock tick tock tick tick tick tick .
- Earth , moon , Earth , moon , full moon , Earth moon .
- crash , boom , crash , boom , crash , boom crash !

Note: **triplicated couple** $A_1B_1A_2B_2A_3B_3$ **always contains an** $A_1B_1B_2A_3$ **implicit antimetabole!!!**

...reversed coordinatives...

- and they splish and they splash and they splash and they splish .
- a dot and a dash and a dash and a dot .
- well Granddad and Grandma [//] Grandma and Granddad are coming today .
- it's called lamb and vegetable [//] mediterranean vegetable and lamb risotto .
- Donald hopped and swam and swam and hopped until he was safe on dry ground .
- every day my cows Poppy (.) Annabel (.) Emily and Heather moo and mumble (.) mumble and moo .
- Chester and Wilson Wilson and Chester .

...and more exhaustive reversed lists...

- Chester and Wilson and Lily Lily and Wilson and Chester .
- okay , square , square , rectangle , square , oval , two , one , one , two .
- blue , green , yellow , red , red , yellow , green , blue .
- one two three or three two one ?
- sure we went through Rhode island , Massachusetts , New Hampshire , Vermont , and then on the way back we did Vermont , New Hampshire , Massachusetts , Rhode island , right ?

...and reversals of direction and position and time...

- you get one ticket that says York to Manchester and another ticket that says Manchester to York .
- he used to rush here and there and there and here and back again all the time and of course he was always in such a rush that he never ever finished anything properly .
- from here to there , from there to here from here to there funny things everywhere .
- let's put mine on yours and put yours on mine .
- could put the box on the lid instead of the lid on the box .
- but I mean do you get your drink after you've had your biscuit or do you get your biscuit after you've had your drink .

...and reversals of attributes...

- let's put the blue one on the guy with the red underpants and the red one on the guy with the blue underpants .
- if it (h)as been a police car it becomes a racing car and if it (h)as been a racing car it becomes a police car .
- and when you're talking about little crocodiles and big snakes (.) or little snakes and big crocodiles (.) they're jelly sweets you've had in the past .
- oh [!] I got a yellow cup and a red plate and you got a red cup and a yellow [!] plate (.) .
- look , they're very similar (.) look , this one is green with a little yellow , and this I yellow with a little green (.) interesting , huh ?
- you mean it looks nicer than it smells [//] smells nicer than it looks .

...and reversals of case-like roles, of course...

Nominative vs. Vocative

- Amanda that's xxx xxx that's Amanda .
- xxx this is Stephanie Stephanie this is xxx by the way .

Nominative vs. Accusative

- froggie keep an eye on mummy or mummy keep an eye on froggie ?
- Floppy meet the screwdrivers screwdrivers meet the Floppy .

Nominative vs. Dative

- do you give Daddy a big kiss or does Daddy give you a big kiss ?

...as well as some more complex swaps?

like Nominative vs. Genitive vs. Locative...

- I mean you go [//] girls go to boys parties and boys go to girls

...or proto-rhetoric questions...

- I think you're stinky you are stinky are you stinky ?
- wouldjou [: would you] couldjou [: could you] wouldjou [: would you] with a goat ?

...and other pieces of maternal wisdom.

- I would not could not in a box I could not would not with a fox .
- we're in house of bricks not the bricks of house .
- two for tea , and tea for two .
- I meant what I said and I said what I meant .

Table of Contents

Concerning the method

- a naive rhetoric-figure-tagger (nRFT)
- fast*, deterministic, transparent for inspection, partially parametrizable
- form-oriented: looks for identic sequences within the signifier (no semantics involved)
- generates false positives: manual check needed; can be useful for $CHIASM_{FP}$ corpus
- can speed-up the manual annotation (semi-supervised scenario)
- IMPORTANT: the schema can be used not only to detect, but also to GENERATE

* and super-fast if You store Your Big Data on a RAMdisk or at least on a SSD disk cache

Concerning the results

- English motherese utterances tend to abound with protochiastic structures
- many functions: playful reversal of repetition, reversal of spatial direction, reversal of list, lapsus lingui correction, positional swap, attribute swap, functional (case) swap ... all matched by a single one-liner !
- what we are dealing here with is a whole **ecosystem** of diverse structures
- indicated prominence of the verb "put" as a middle term consistent with theories of Piaget and Tomasello
- triplicated couple $A_1 B_1 A_2 B_2 A_3 B_3$ always contains an $A_1 B_1 B_2 A_3$ implicit antimetabole

Invitation to explore

- not only intralocutory (i.e. within 1 utterance) chiasms, but also translocutory ones (within multiple successive utterances)
- relations to variation sets and Winograd schemata
- multi-lingual analysis (are these beasts universal ?)
- ontogenetic relation to other figures like rhetoric question or even metaphore (METAPHOROS = "carry over")
- informational content of chiasms (known components + unknown order = maximal amount of new info ?)
- neurocognitive aspects of chiasm processing (focus upon the **cyclical referential closure** between initial and terminal token of the sequence)
- neurorhetoric hypothesis: look for a P600-like evoked potential following the exposure to chiasmus
- non-linguistic chiasmata (musical, visual, spatial, anatomical, social, moral, emotional, sexual, spiritual etc.)

Conclusion

Starting discussion with conclusion often concludes the discussion...

Ergo, no ultimate conclusion without juicy discussion.

daniel@wizzion.com thanks Thee for Thy attention