Motivation

Strict distinction between surface realization and sentence planning may cause difficulties for an NLG system:

- Generation decisions made by the sentence planner may be hard or impossible to realize.
- Computing each referring expression (RE) separately might cause the planner to miss out opportunities for succinct REs that are ambiguous in isolation but correct in context.

Our proposal

A chart-based algorithm for integrated referring expression generation and surface realization of full sentences.

Semantically interpreted grammars (SIG)

An example SIG template:

for all e, a ∈ sleep:
S_a → sleep_p_a(NP_e)
I_a(sleep, p_a)(w) = w \{ sleep \}
I_a(sleep, p_a)(R) = \{ sleep \} \cap uniq(R)
I_a(sleep, p_a)(N) = \{ sleep(e, a) \} = N

for all a ∈ rabbit:
N_a → rabbit_p_a(Adj)
I_a(rabbit, p_a)(w) = w \{ rabbit \}
I_a(rabbit, p_a)(R) = \{ rabbit \} \cap R
I_a(rabbit, p_a)(N) = \{ rabbit(e, a) \} = N

for all a ∈ U:
NP_a → def_a(N)
I_a(def)(w) = \{ w \}
I_a(def)(R) = R
I_a(def)(N) = N

Each derivation tree of a SIG is interpreted as:

- a set of references (I_R interpretation)
- a set of semantic atoms (I_N interpretation)
- a string (I_S interpretation)

Integrated sentence generation

The chart computation algorithm:

Excerpt from the chart for “the white rabbit sleeps”:

| Target semantics: N = \{ takefrom(e_2, r_1, h_1) \} | Target referent: R = \{ e_2 \} |
|-----------------------------------------------|
| “Take the rabbit from the hat.” |

Generating contextually unique REs

“Take the rabbit from the hat”

- Assume targets R = \{ e_2 \} and N = \{ takefrom(e_2, r_1, h_1) \}.
- The chart algorithm will construct: 1. \( t_1 = \text{def}_1(\text{rabbit}(\text{nop}_1)) \) (“the rabbit”) 2. \( t_2 = \text{def}_2(\text{hat}(\text{nop}_1)) \) (“the hat”) with \( R_1 = I_{R_1}(t_1) = \{ r_1, t_2 \} \) and \( R_2 = I_{R_2}(t_2) = \{ h_1, h_2 \} \).
- \( R_1 \cap_1 \{ \text{in} \cap_2 R_2 \} \) evaluates to \( \{ r_1 \} \), satisfying uniqueness.
- \( t = \text{takefrom}(e_2, n, r_1, h_2) \) is a valid realization.

We let REs mutually constrain each other by moving the requirement for semantic uniqueness to the verb.

Chart computation heuristics

Surface realization, generation of shortest REs: NP-complete

⇒ Use heuristics to avoid computing the whole chart.

Heuristic for generating sentences with shortest REs:

1. Compute full chart \( C_R \) for the \( I_R \) part of the input.
2. Compute the distance of each chart item to a goal item.
3. Add the \( I_R \) parts to each chart item.
4. Order unprocessed items \([A, R, N]\) by:
   a) #atoms left unrealized in the target semantic content
   b) distance of \([A, R]\) to a goal item in \( C_R \).
5. Stop chart computation once the first goal item is found.

Conclusion

- A chart-based algorithm for integrated referring expression generation and surface realization of full sentences.
- Agnostic on the grammar formalism on the string side – e.g. can use a TAG instead of a CFG.
- Structure-sharing and pruning techniques could help use integrated sentence generation in practical applications.
- Open-source code can be found at bitbucket.org/tclup/altog