Lexicon:

uywon senator

kica reporter

sasil fact

ul -ACC

ka -NOM

i -NOM

kongkyekhan attack.ADN

yumyenghaycyessta became famous

palkhyecyessta was.revealed

Global Renormalization

**Korean Relative Clauses**

Entropy Reductions:

uywon/kica ul/ka kongkyekhan kica/uywon ka/i yumyenghaycyessta

SRC 0.9747775695 0 0.365613444 0 6.058694209 4.2814254193

ORC 0.9747775695 4.5261632779 0 0.975830001 6.058694209 4.2814254193

Entropy:

SRC 8.0360411735 7.0612636039 10.1895237201 9.8239102760 10.3401196288 4.2814254193 0.\*10^-14

ORC 8.0360411735 7.0612636039 2.5351003260 11.3159496301 10.3401196288 4.2814254193 0.\*10^-14

Results:

The SRC advantage is observed at as early as the second word in the sentence (i.e. ul in the SRC; ka in the ORC)

The ERs at the head noun (kick in the SRC; uywon in the ORC) also suggest a slight SRC preference.

There is no difference of ERs at the main verb.

**Question 1: why is the conditional entropy so much higher in case where a sentence begins with an ACC NP than a NOM NP?**

"uywon ul" (reporter-NOM) is the prefix that starts the SRC sentence.

after "ul", entropy is at 10.190 bits, before "ul" it is at 7.061 bits. (i.e. ER = 0)

"kica ka" (reporter-NOM) is the prefix that starts the ORC sentence.

after "ka", entropy is at 2.535 bits, before "ka" it is at 7.061 bits.(i.e. ER = 4.53)

The greater reduction of uncertainty from "kica" to "kica ka" comes from the two dominating main clauses with intransitive verbs after having the NOM marker. A sentence-initial nominative NP will is more likely to lead a short matrix clause.

Top 5 strings are

1. 0.516305934893 “kica ka yumyenghaycyessta”

2. 0.37253350394 “kica ka palkhyecyessta”

3. 0.0131435045802 “kica ka uywon ul kongkyekhayssta”

4. 0.00747574916294 “kica ka kica ul kongkyekhayssta”

5. 0.00517197228193 “kica ka kongkyekhan uywon ka yumyenghaycyessta” (\*uywon ka is morphologically incorrect\*)

whereas given the prefix "uywon ul", strings are relatively equiprobable just as derivations given "uywon". There is no "simple" way to complete the sentence with just a single verb when having a sentence initial accusative NP. Completing the sentence with a transitive verb will result in a subject pro-drop main clause and only ranks at #48.

1. 0.0725650262397 “uywon ul kongkyekhan sasil ka yumyenghaycyessta”

2. 0.0658606488726 “uywon ul kongkyekhan sasil i yumyenghaycyessta”

3. 0.0523583047601 “uywon ul kongkyekhan sasil ka palkhyecyessta”

4. 0.0475208527311 “uywon ul kongkyekhan sasil i palkhyecyessta”

5. 0.0252969389915 “uywon ul kongkyekhayse uywon ka yumyenghaycyessta”

……

48. 0.00173870086503 “uywon ul kongkyekhayssta”

**Question 2: what happened at the third word "kongkyekhan", an adnominal-marked verb, i.e. ER= 0.366 bits in SRC and 0 in ORC**

In SRC

The "uywon ul" has three possible continuations indicating four different clause types.

MC: uywon ul kongkyekhayssta

AC: uywon ul kongkyekhayese

CC: uywon ul kongkyekhan sasil

SRC: uywon ul kongkyekhan kica

In top ten strings, CC (rank 1-4) outranks AC (rank 5-6,9-10) and SRC (rank 7-8)

……

6. 0.022959721822 “uywon ul kongkyekhayse uywon i yumyenghaycyessta”

7. 0.0221220517119 “uywon ul kongkyekhan uywon ka yumyenghaycyessta”

8. 0.0200781665168 “uywon ul kongkyekhan uywon i yumyenghaycyessta”

9. 0.0182526612316 “uywon ul kongkyekhayse uywon ka palkhyecyessta”

10. 0.0165662740669 “uywon ul kongkyekhayse uywon i palkhyecyessta”

"uywon ul kongkyekhan" rules out the possibility of MC and AC. CC and SRC are the two possible continuations remaining.

In ORC

There is a huge entropy gain (8.78 bits) from "kica ka" to "kica ka kongkyekhan".

Recall in "kica ka", two dominating structures are main clauses with intransitive verbs.

The verb.ADN "kongkyekhan" suggest the continuation to be either a CC or an ORC.

No dominating structure in this case. Probabilities mass has been diluted.

Top five strings are ORC

1. 0.0700400689177 “kica ka kongkyekhan uywon ka yumyenghaycyessta”

2. 0.063568975649 “kica ka kongkyekhan uywon i yumyenghaycyessta”

3. 0.0505364562496 “kica ka kongkyekhan uywon ka palkhyecyessta”

4. 0.0458673271794 “kica ka kongkyekhan uywon i palkhyecyessta”

5. 0.0398373191405 “kica ka kongkyekhan kica ka yumyenghaycyessta”

However, as we saw from 100 top strings. A large number of generated structure involves different kinds of multiple embedding. (\*need to find out looping during generation!\*)

Some of them are infelicitous. For example, a improper double-embedded complement clause

13. 0.00179851434964 “[[kica ka kongkyekhan uywon ul kongkyekhan] sasil ul kongkyekhan] sasil i yumyenghaycyessta”

[the fact that [the fact that the reporter attacked the senator] attacked someone] became famous

To be short, "kica ka kongkyekhan" is a prefix that could leaves to a great uncertainty.

**Question 3: why there is a small SRC preference at the head noun?**

In ORC, adding the next noun kica leads to a mild disambiguation because it rules out a number of embedded sentences.

31. 7.81826148009e-6 “kica ka kongkyekhan kongkyekhan sasil ul kongkyekhan kica ka yumyenghaycyessta”

The possibility mass moves towards to top derivations.

1. 0.11031384364 “kica ka kongkyekhan uywon ka yumyenghaycyessta”

2. 0.100121803825 “kica ka kongkyekhan uywon i yumyenghaycyessta”

3. 0.0795954489904 “kica ka kongkyekhan uywon ka palkhyecyessta”

4. 0.0722415217007 “kica ka kongkyekhan uywon i palkhyecyessta”

**Korean Complement Clauses**

uywon/kica ul/ka kongkyekhan sasil i palkhyecyessta

SCC 0.9747775695 0 0.365613444 2.7626466720 4.526163277 2.5351003260

OCC 0.9747775695 4.5261632779 0 5.1048654661 1.4875475232 4.7235366407

SCC 8.0360411735, 7.0612636039, 10.1895237201, 9.8239102760, 7.0612636039, 2.5351003260, 0.\*10^-14

OCC 8.0360411735, 7.0612636039, 2.53510032601, 11.3159496301, 6.2110841640, 4.7235366407, 0.\*10^-14

The first three words are the same with relative clauses, i.e. uywon/kica ul/ka kongkyekhan.

We see a large subject preference at the second set of word NOM/ACC and a slight object preference at the embedded verb.

**Question 4: the subject preference at the head noun is larger than we saw in RC conditions.**

In SCC, the entropy before “sasil” is 9.8 bits, after “sasil” 7.1 bits. Apparently, the top four derivations gain more probability mass, which reduces the uncertainty a little bit.

before “sasil”

1. 0.0965374517743 “uywon ul kongkyekhan sasil ka yumyenghaycyessta”

2. 0.0876182307626 “uywon ul kongkyekhan sasil i yumyenghaycyessta”

3. 0.0696552813757 “uywon ul kongkyekhan sasil ka palkhyecyessta”

4. 0.0632197391296 “uywon ul kongkyekhan sasil i palkhyecyessta”

5. 0.0294302449947 “uywon ul kongkyekhan uywon ka yumyenghaycyessta”

after “sasil”

1. 0.166314879166 “uywon ul kongkyekhan sasil ka yumyenghaycyessta”

2. 0.150948830678 “uywon ul kongkyekhan sasil i yumyenghaycyessta”

3. 0.120002232215 “uywon ul kongkyekhan sasil ka palkhyecyessta”

4. 0.108915069551 “uywon ul kongkyekhan sasil i palkhyecyessta”

5. 0.0279747010338 “uywon ul kongkyekhan sasil ul kongkyekhan sasil ka yumyenghaycyessta”

In OCC, the entropy before “sasil” is 11.3 bits, after “sasil” 6.2 bits. Recall that the prefix string “kica ka kongkyekhan” poses a huge uncertainty because of multiple embedded structures. After “sasil”, we see a dominating continuation with a 0.44 probability, although semantically strange? (“Someone attacked the fact that the reporter attacked someone”)

before “sasil”

1. 0.0700400689177 “kica ka kongkyekhan uywon ka yumyenghaycyessta”

2. 0.063568975649 “kica ka kongkyekhan uywon i yumyenghaycyessta”

3. 0.0505364562496 “kica ka kongkyekhan uywon ka palkhyecyessta”

4. 0.0458673271794 “kica ka kongkyekhan uywon i palkhyecyessta”

5. 0.0398373191405 “kica ka kongkyekhan kica ka yumyenghaycyessta”

after “sasil”

1. 0.446261461141 “kica ka kongkyekhan sasil ul kongkyekhayssta”

2. 0.0750626223037 “kica ka kongkyekhan sasil ul kongkyekhan sasil ul kongkyekhayssta”

3. 0.0371969409518 “kica ka kongkyekhan sasil ka yumyenghaycyessta”

4. 0.0337602670888 “kica ka kongkyekhan sasil i yumyenghaycyessta”

5. 0.0268389453076 “kica ka kongkyekhan sasil ka palkhyecyessta”

**Question 5: the subject-object asymmetry pattern reverses at the marker for the head noun in a complement clause**

In SCC, before “i” 7.1 bits, after “i” 2.5 bits; the large reduction comes from two no-longer available and \*infelicitous\* derivation. (i.e. with a head noun “sasil ka”) Note: it seems that morphology does influence the results!

before “i”

1. 0.166314879166 “uywon ul kongkyekhan sasil ka yumyenghaycyessta”

2. 0.150948830678 “uywon ul kongkyekhan sasil i yumyenghaycyessta”

3. 0.120002232215 “uywon ul kongkyekhan sasil ka palkhyecyessta”

4. 0.108915069551 “uywon ul kongkyekhan sasil i palkhyecyessta”

…

after “i”

1. 0.516305934893 “uywon ul kongkyekhan sasil i yumyenghaycyessta”

2. 0.37253350394 “uywon ul kongkyekhan sasil i palkhyecyessta”

3. 0.0131435045802 “uywon ul kongkyekhan sasil i uywon ul kongkyekhayssta”

4. 0.00747574916294 “uywon ul kongkyekhan sasil i kica ul kongkyekhayssta”

5. 0.00517197228193 “uywon ul kongkyekhan sasil i kongkyekhan uywon ka yumyenghaycyessta”

In OCC, before “i” 6.2 bits, after “i” 4.7 bits.

before “i”

1. 0.446261461141 “kica ka kongkyekhan sasil ul kongkyekhayssta”

2. 0.0750626223037 “kica ka kongkyekhan sasil ul kongkyekhan sasil ul kongkyekhayssta”

3. 0.0371969409518 “kica ka kongkyekhan sasil ka yumyenghaycyessta”

4. 0.0337602670888 “kica ka kongkyekhan sasil i yumyenghaycyessta”

5. 0.0268389453076 “kica ka kongkyekhan sasil ka palkhyecyessta”

after “i”

1. 0.408102219881 “kica ka kongkyekhan sasil i yumyenghaycyessta”

2. 0.294460589475 “kica ka kongkyekhan sasil i palkhyecyessta”

3. 0.0490456153914 “kica ka kongkyekhan sasil i kongkyekhan uywon ul kongkyekhayssta”

4. 0.0278961152235 “kica ka kongkyekhan sasil i kongkyekhan kica ul kongkyekhayssta”

5. 0.0103889826432 “kica ka kongkyekhan sasil i uywon ul kongkyekhayssta”

***Korean Surprisal***

SRC 0.69703267643, 1.37515167888, 0.41181435119, 2.72030234495, 2.06936389777, 1.11095032258

ORC 1.51109262237, 1.63430892825, 3.75939405856, 0.65536145405, 2.20922156136, 1.11095032258

SCC 0.69703267643, 1.37515167888, 0.41181435119, 0.78475660216, 1.77416659183, 1.42455792348

OCC 1.51109262237, 1.63430892825, 3.75939405856, 9.35554647089, 3.59553234077, 1.76385354809

**Koreanabs Report**

***Entropy Reduction:***

Noun Acc/Nom Vadn Noun Nom/Acc Vdecl

SRC 0.037175732 0 0.24631722 0 3.44026814 3.095744963

ORC 0.037175732 2.371901811 0 0.07276697 3.44026814 3.095744963

Noun Acc/Nom Vadn fact Nom/Acc Vdecl

SCC 0.037175732 0 0.24631722 1.73608079 2.371901811 2.381236873

OCC 0.037175732 2.371901811 0 4.639960235 0 4.282586545