Issues in Korean/English Machine Translation

Chung-hye Han, Nari Kim, Martha Palmer

University of Pennsylvania
Feb. 2-4, 1999

Word Order and Scrambling

source	추가 공급물을 103 전위지원대대에게 사령부가 주었다. (chwuka kongkupmwulul 103 cenwiciwontaytayeykey salyengpwuka cwuessta.)
Glosser	additional supply$_2$ 103 FSB$_1$ headquarter$_0$ gave
OTS MT system	Additional supply$_2$ 103 FSB$_1$ headquarters$_0$ which you bite $\boxed{\text{gave}}$.
target	Headquarters$_0$ $\boxed{\text{gave}}$ 103rd FSB$_1$ additional supplies$_2$.

Empty Arguments

source	우리의 최근 요청을 수신했는가? (wuliuy choykun yochengul swusinhayssnunka?)
Glosser	our current request received
OTS MT system	..... To receive an our recent request?
target	Did (NP) receive our current request?

Empty Arguments

source	IBP 화일을 검색 못했다면 지금 다시 보내겠다. (IBP hwailul kemsayk moshayesstamyen cikum tasi ponaykeyssta.)
Glosser	IBP file retrieval could not now again will send
OTS MT system	The IBP the file the retrieval the cotton now not to be able you will send again.
target	If (NP1) could not retrieve IBP file, (NP2) will send again now.

Lexical Ambiguity

- Verb: **lose**

  잃어버리다 (ilepelita): e.g., lose$_1$ an object

  지다 (cita): e.g., lose$_2$ a battle

source	Iraq lost the battle.
target	이라크가 전투에 졌다. (ilakuka centwuey **ciessta**.) 'Iraq **lost** the battle.'
wrong target	*이라크가 전투를 잃어버렸다. (*ilakuka centwulul **ilepelyessta**.) 'Iraq **misplaced** the battle.'

- Postposition: **-으로** (**-ulo**)

  by, as, to, with

source	판단서는 서면-**으로** 제출되었다. (phantansenun semyen-**ulo** ceychwultoyessta.)
target	Estimates were presented **as** a written document.
OTS MT system	*The estimates presented **with** the written document.

source	부대는 전방-**으로** 이동했다. (pwutaynun cenpang-**ulo** itonghayssta.)
target	The unit moved **to** the front line.
OTS MT system	*The squad moved **with** the front line.

Issues in Korean/English Machine Translation

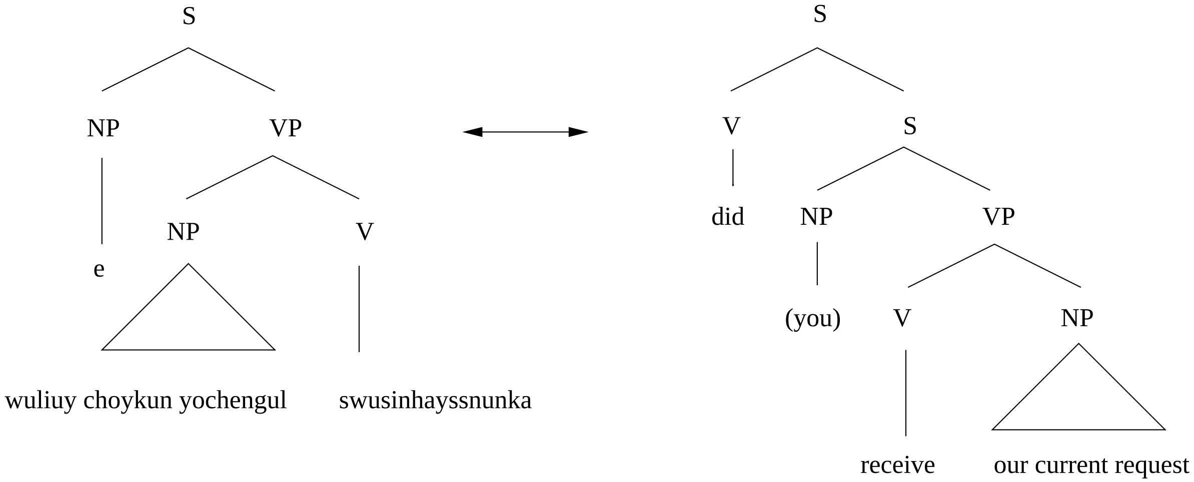
Chung-hye Han, Nari Kim, Martha Palmer

University of Pennsylvania

Feb. 2-4, 1999

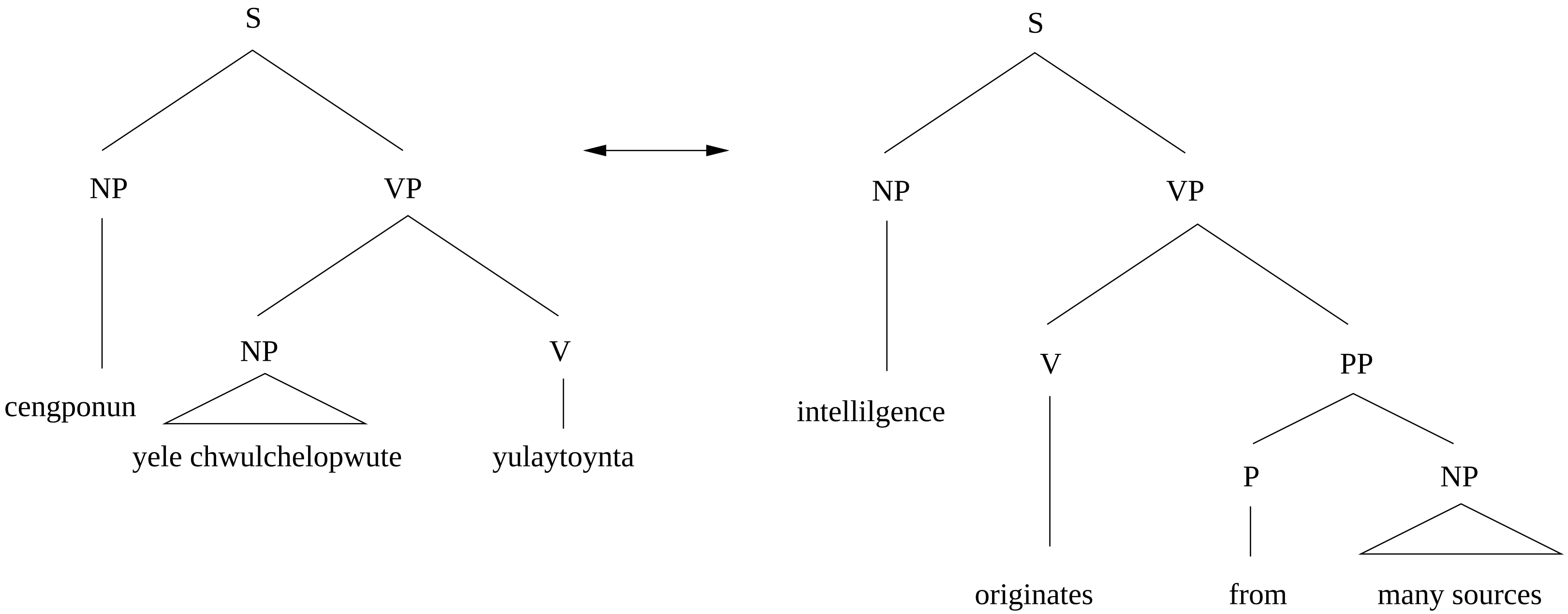
Word Order

source	지휘관은 정보활동을 지시한다. (cihwikwanun cengpohwaltongul cisihanta.)
Glosser	commander intelligence activity instruct
OTS MT system	The commander instructs intelligence activity.
target	The commander instructs intelligence activity.


S
NP
VP
e
NP
V
wuliuy choykun yochengul
swusinhayssnunka
S
V
S
did
NP
VP
(you)
V
NP
receive
our current request

Prepositions vs. Postpositions

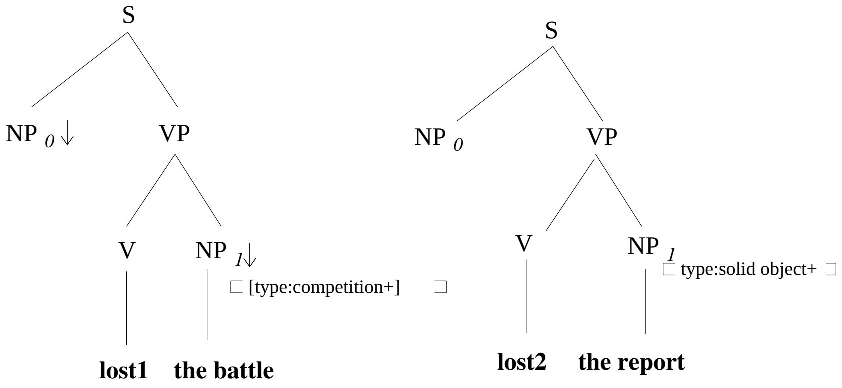
source	정보는 여러 출처로부터 유래된다. (cengponun yele chwulchelopwute yulaytoynta.)
Glosser	intelligence many source originate
OTS MT system	Intelligence is derived from the multi source.
target	Intelligence originates from many sources.


S
NP
VP
NP
V
cengponun
yele chwulchelopwute
yulaytoynta
S
NP
VP
V
PP
P
NP
intellilgence
originates
from
many sources

Word Order and Scrambling

source	추가 공급물을 103 전위지원대대에게 사령부가 주었다. (chwuka kongkupmwulul 103 cenwiciwontaytayeykey salyengpwuka cwuessta.)
Glosser	additional supply 103 FSB headquarter gave
OTS MT system	Additional supply 103 FSB headquarters which you bite gave.
target	Headquarters gave 103rd FSB additional supplies.

Selectional Restrictions for Lexical Choice



lost1 -> ciessta
lost2 -> ilepeliessta

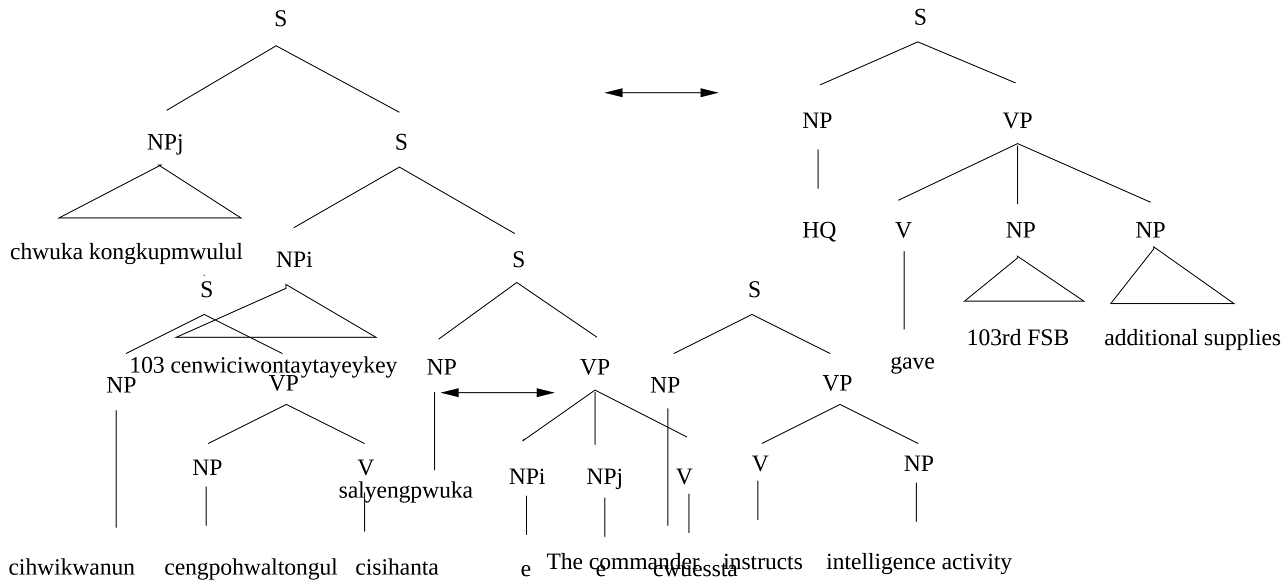
Empty Arguments

source	우리의 최근 요청을 수신했다. (wuliuy choykun yochengul swusinhayssta.)
Glosser	our current request received
OTS MT system	It receives an our recent request.
target	(NP) received our recent request.

source	우리의 최근 요청을 수신했는가? (wuliuy choykun yochengul swusinhayssnunka?)
Glosser	our current request received
OTS MT system	To receive an our recent request?
target	Did (NP) receive our current request?

Empty Arguments

source	IBP 화일을 검색 못했다면 지금 다시 보내겠다. (IBP hwailul kemsayk moshayesstamyen cikum tasi ponaykeyssta.)
Glosser	IBP file retrieval could not now again will send
OTS MT system	The IBP the file the retrieval the cotton now not to be able you will send again.
target	If (NP1) could not retrieve IBP file, (NP2) will send again now.


S
NPj
S
chwuka kongkupmwulul
NPi
S
S
103 cenwiciwontaytayeykey
NP
VP
NP
VP
NP
V
salyengpwuka
NPi
NPj
V
cihwikwanun
cengpohwaltongul
cisihanta
e
e
cwuessta
S
NP
VP
HQ
V
NP
NP
gave
103rd FSB
additional supplies
S
NP
VP
V
NP
The commander
instructs
intelligence activity

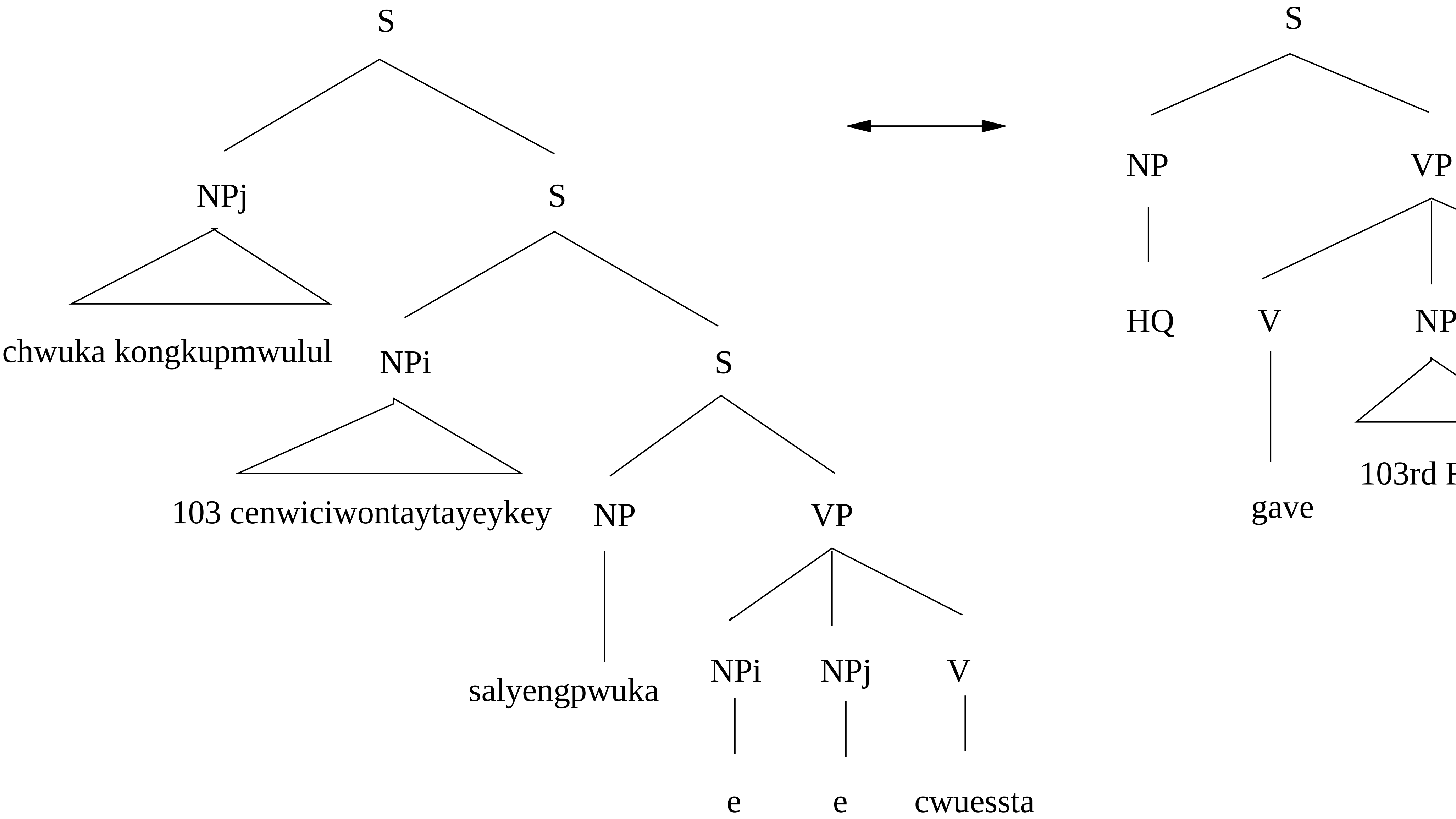
Lexical Ambiguity

- Verb: **lose**

  잃어버리다 (ilepelita): e.g., lose_1 an object

  지다 (cita): e.g., lose_2 a battle

source	Iraq lost the battle.
target	이라크가 전투에 졌다. (ilakuka centwuey ciessta.) ‘Iraq **lost** the battle.’
wrong target	*이라크가 전투를 잃어버렸다. (*ilakuka centwulul ilepelyessta.) ‘Iraq **misplaced** the battle.’


S
NPj
S
chwuka kongkupmwulul
NPi
S
103 cenwiciwontaytayeykey
NP
VP
salyengpwuka
NPi
NPj
V
e
e
cwuessta
S
NP
VP
HQ
V
NP
gave
103rd

- Postposition: **-으로 (-ulo)**

  by, as, to, with

source	판단서는 서면-**으로** 제출되었다. (phantansenun semyen-**ulo** ceychwultoyessta.)
target	Estimates were presented **as** a written document.
OTS MT system	*The estimates presented **with** the written document.
source	부대는 전방-**으로** 이동했다. (pwutaynun cenpang-**ulo** itonghayssta.)
target	The unit moved **to** the front line.
OTS MT system	*The squad moved **with** the front line.