

OntoNotes Coreference

2009-02-24

Lance Ramshaw & Michelle Franchini









Coref Annotation Examples



IDENT

- [Elco Industries Inc.] $_{\rm x}$ said [it] $_{\rm x}$ expects net income to fall below a recent estimate of \$ 1.65 a share. [The Rockford, III. maker of fasteners] $_{\rm x}$ also said [it] $_{\rm x}$ expects to post sales in the current fiscal year that are "slightly above" fiscal 1989 sales of \$ 155 million.
- Sales of passenger cars [grew]_x 22%. [The strong growth]_x followed year-to-year increases.

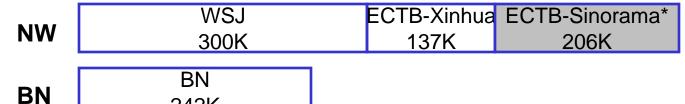
APPOS

[[the PhacoFlex intraocular lens]<HEAD>, [the first foldable silicone lens available for cataract surgery]<ATTRIB>]

Completed Corpora



English



BC ENG ENG-fr-CHI 179K 70K

242K

Chinese



BN	BN**
	297K



Arabic

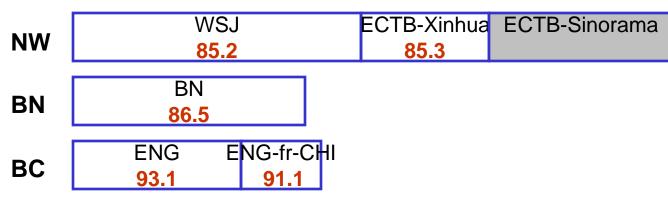
NW	ATB	ATB
	100K	100K

- * Double annotated but not yet fully adjudicated
- ** Not yet fully double annotated or adjudicated

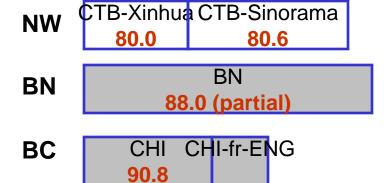
Agreement Rates (mean of Ann1:Adj and Ann2:Adj)



English



Chinese



Arabic

NW	ATB	ATB
	75.4	83.1

Special Challenges in the BC Data



Disfluency effects:

Former Iraqi war combat veteran I guess 0 [disfluency] he 's a -- -- [/disfluency] he is a present veteran Paul Hackett ...

Ambiguity in speaker turn labels:

[Firefighter_A] It began as [disfluency] an- [/disfluency] any other day you know *PRO* [uncertain] just uh [/uncertain] doing eh normal checks.

[Firefighter_B] At nine o'clock we started our shift .

[Firefighter_A] And so the bells went .

[Firefighter_B] It was about a minute past nine when we got the shout for uh [uncertain] smoke issuing [/uncertain] in Allgate tube station *T*-1.

[Andrew_Carey] The explosion at Allgate was the first of the four bombs 0 *T*-1 to go off on July the seventh at eight fifty in the morning.

[Andrew_Carey] But Paul Kelly Steve Sodbury and Mel Anderson of [uncertain] Shadwell [/uncertain] Firestation 's blue watch had no idea what *T*-1 had happened as they got into the fire engine *PRO*-2 to answer the call .