Stanford Dependencies (SDs)

Instructions

1. Have a look over the tagging manual, specifically Sections 1 and 2.

    - We are doing the **basic** style, briefly described in 4.1. In short - basic dependencies are trees (fully connected, no cycles), and not all the dependencies in Section 2 apply - those that don't have a note in their description saying that. List of such dependencies

    - A **downloadable example text file** (all labeled by linguists) can be found here. (I'm not sure which exact version of SDs these are so be aware they may not be the basic variant. A comparison of the different versions labels is given in Section 4.8 so it's easy to convert.)

    - For those who like **graphic examples** like me, I rendered the first 100 trees in the PTB as basic SDs here (sorry that the formatting is crude and some get a bit tiny). You can clearly see the differences from GRs (e.g. look how *conj* and *cc* is done in sentence 2, or how punctuation is done in sentence 15). It's searchable so you can ctrl+f a particular word or tag.
        i. ***Link updated 12.45pm 27/12** - now shows new dependencies from the official release of WSJ PTB (be aware all sentences are changed)*
        ii. *Added a conll upload with many more examples.*

    - Under *punct* (p.9), it says "by default, punctuation is not retained in the output". However, I think it would be useful to label punct, since it means that all tokens are connected to exactly one other token. Besides, it's really trivial.

2. Fill in your sentences, according to the assignments on the next page. **Please do so by the end of the 28th Dec!!!**

    - Most of the sentences already have GR annotation in the old google doc, so refer to that along with the comments. You may wish to be reminded of the GR tagging manual.

    - For the ones with no GR annotation - start from scratch, write out your answer, THEN compare with the other person who is annotating the same sentence. **Then communicate with the other person and resolve differences, and together fill out a gold standard.**

Format

For the format I think we should use **CoNLL-X** because then they are easily machine readable, standard for SDs, and for our purposes it is identical to CoNLL-U.

Here is a good very quick intro explaining what is in each tab-separated column. **We are only interested in columns 1, 2, 7, and 8**, so all other columns should just be "_". Below is a pasted example from the PTB (note it has more filled in than we need):

```
1       Ms.     _       NNP     NNP     _       2       nn      _       _
2       Haag    _       NNP     NNP     _       3       nsubj   _       _
3       plays   _       VBZ     VBZ     _       0       root    _       _
4       Elianti _       NNP     NNP     _       3       dobj    _       _
5       .       _       .       .       _       3       punct   _       _
```

Tokens are indexed from 1; the root is index 0. Use tabs, don't use spaces - otherwise it won't be machine readable.

Assignments

Bold indicates that there is no current annotation present in the GR doc. Everyone gets one bold one. I gave the excess to the people who had easier bold ones. I've tried to do it fairly but let me know if you want to swap around!

Jennifer	2	**11**	20	7	12	
Martyna	3	**13**	21	8	14	
Lucas	**4**	12	22	9	15	
Rowan	**5**	14	23		1	
Roddy	**6**	15	24	10		
Stefan	7	16	25	**5**		
Irene	8	17	1	**6**	18	
Ines	9	18	2	**11**	19	
Simone	10	19	3	**13**	20	
Rami	21	22	23	24		
Yifan	25	17	16	**4**		
Chen (Kuan-Chi)	2	7	9	18	20	25

Daan	Check comments	3	12

Put down your name before your annotation. At the end, all will have 2 annotators.

I filled out the first one as an example of what it should look like, format-wise.

Done list:

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

2ov0 21 22 23 24 25

Sentences

(1) The old car broke down in the car park.

Agreed by Rowan and Irene and Daan:

```
1     The    _     _     _     _     3     det    _     _
2     old    _     _     _     _     3     amod   _     _
3     car    _     _     _     _     4     nsubj  _     _
4     broke  _     _     _     _     0     root   _     _
5     down   _     _     _     _     4     prt    _     _
6     in     _     _     _     _     4     prep   _     _
7     the    _     _     _     _     9     det    _     _
8     car    _     _     _     _     9     nn     _     _
9     park   _     _     _     _     6     pobj   _     _
10    .      _     _     _     _     4     punct  _     _
```

(2) At least two men broke in and stole my TV.

Jennifer (+ checked over by Rowan,Yifan, Kuan-Chi, Ines):

```
1     At     _     _     _     _     3     quantmod   _     _
2     least  _     _     _     _     1     mwe        _     _
3     two    _     _     _     _     4     num        _     _
4     men    _     _     _     _     5     nsubj      _     _
5     broke  _     _     _     _     0     root       _     _
6     in     _     _     _     _     5     prt        _     _
7     and    _     _     _     _     5     cc         _     _
8     stole  _     _     _     _     5     conj       _     _
9     my     _     _     _     _     10    poss       _     _
10    TV     _     _     _     _     8     dobj       _     _
11    .      _     _     _     _     5     punct      _     _
```

(3) The horses were broken in and ridden in two weeks.

Daan (+ checked by Jennifer, Lucas):

```
1     The     _     _     _     _     2     det     _     _
2     horses  _     _     _     _     4     nsubjpass  _     _
3     were    _     _     _     _     4     auxpass    _     _
4     broken  _     _     _     _     0     root    _     _
5     in      _     _     _     _     4     prt     _     _
6     and     _     _     _     _     4     cc      _     _
7     ridden  _     _     _     _     4     conj    _     _
8     in      _     _     _     _     4     prep    _     _
9     two     _     _     _     _     10    num     _     _
10    weeks   _     _     _     _     8     pobj    _     _
11    .       _     _     _     _     4     punct   _     _
```

(4) Kim and Sandy both broke up with their partners.

```
# Yifan + Lucas (+ Roddy):
# Discussion about “both” here. Tl;dr: See CoNLL 46960.
1     Kim      _     _     _     _     5     nsubj _     _
2     and      _     _     _     _     1     cc    _     _
3     Sandy    _     _     _     _     1     conj  _     _
4     both     _     _     _     _     3     det   _     _
5     broke    _     _     _     _     0     root  _     _
6     up       _     _     _     _     5     prt   _     _
7     with     _     _     _     _     5     prep  _     _
8     their    _     _     _     _     9     poss  _     _
9     partners _     _     _     _     7     pobj  _     _
10    .        _     _     _     _     5     punct _     _
```

(5) The horse which Kim sometimes rides is more bad tempered than mine.

```
# Rowan + Stefan (+Irene, Roddy, Daan):
# Length discussion about ‘bad’s label here
1     The        _     _     _     _     2     det    _     _
2     horse      _     _     _     _     10    nsubj  _     _
3     which      _     _     _     _     6     dobj   _     _
4     Kim        _     _     _     _     6     nsubj  _     _
5     sometimes  _     _     _     _     6     advmod       _     _
6     rides      _     _     _     _     2     rcmod  _     _
7     is         _     _     _     _     10    cop    _     _
8     more       _     _     _     _     10    advmod       _     _
9     bad        _     _     _     _     10    advmod       _     _
10    tempered   _     _     _     _     0     root   _     _
11    than       _     _     _     _     10    prep   _     _
12    mine       _     _     _     _     11    pobj   _     _
13    .          _     _     _     _     10    punct  _     _
```

(6) The horse as well as the rabbits which we wanted to eat have escaped.

```
# Irene + Roddy (+ Daan, Lucas, Jennifer):
# See 2  for discussion on “wanted”
1     The    _     _     _     _     2     det    _     _
2     horse  _     _     _     _     14    nsubj  _     _
3     as     _     _     _     _     4     mwe    _     _
4     well   _     _     _     _     2     cc     _     _
5     as     _     _     _     _     4     mwe    _     _
6     the    _     _     _     _     7     det    _     _
7     rabbits_     _     _     _     2     conj   _     _
8     which  _     _     _     _     10    dobj   _     _
9     we     _     _     _     _     10    nsubj  _     _
10    wanted_      _     _     _     7     rcmod  _     _
11    to     _     _     _     _     12    aux    _     _
12    eat    _     _     _     _     10    xcomp  _     _
13    have   _     _     _     _     14    aux    _     _
14    escaped_     _     _     _     0     root   _     _
15    .      _     _     _     _     14    punct  _     _
```

(7) It was my aunt's car which we sold at auction last year in February.

Jennifer (+ checked over by rowan, Yifan, Kuan-Chi):

```
1     It       _     _     _     _     6     nsubj    _     _
2     was      _     _     _     _     6     cop      _     _
3     my       _     _     _     _     4     poss     _     _
4     aunt     _     _     _     _     6     poss     _     _
5     's       _     _     _     _     4     possessive _     _
6     car      _     _     _     _     0     root     _     _
7     which    _     _     _     _     9     dobj     _     _
8     we       _     _     _     _     9     nsubj    _     _
9     sold     _     _     _     _     6     rcmod    _     _
10    at       _     _     _     _     9     prep     _     _
11    auction  _     _     _     _     10    pobj     _     _
12    last     _     _     _     _     13    amod     _     _
13    year     _     _     _     _     9     tmod     _     _
14    in       _     _     _     _     9     prep     _     _
15    February _     _     _     _     14    pobj     _     _
16    .        _     _     _     _     6     punct    _     _
```

(8) The only rabbit that I ever liked was eaten by my parents one summer.

Irene (+Jennifer + Rowan)

1	The	_	_	_	_	3	det	_	_
2	only	_	_	_	_	3	amod	_	_
3	rabbit	_	_	_	_	9	nsubjpass	_	_
4	that	_	_	_	_	7	dobj	_	_
5	I	_	_	_	_	7	nsubj	_	_
6	ever	_	_	_	_	7	advmod	_	_
7	liked	_	_	_	_	3	rcmod	_	_
8	was	_	_	_	_	9	auxpass	_	_
9	eaten	_	_	_	_	0	root	_	_
10	by	_	_	_	_	12	prep	_	_
11	my	_	_	_	_	12	det	_	_
12	parents	_	_	_	_	10	pobj	_	_
13	one	_	_	_	_	14	num	_	_
14	summer	_	_	_	_	9	tmod	_	_
15	.	_	_	_	_	9	punct	_	_

(9) The veterans who I thought that we would meet at the reunion were dead.

Ines + Lucas (+ Daan + Jennifer + Kuan-Chi)
See here for comment on the head of “who”

```
1     The       _     _     _     _     2     det     _     _
2     veterans  _     _     _     _     14    nsubj   _     _
3     who       _     _     _     _     9     dobj    _     _
4     I         _     _     _     _     5     nsubj   _     _
5     thought   _     _     _     _     2     rcmod   _     _
6     that      _     _     _     _     9     mark    _     _
7     we        _     _     _     _     9     nsubj   _     _
8     would     _     _     _     _     9     aux     _     _
9     meet      _     _     _     _     5     ccomp   _     _
10    at        _     _     _     _     9     prep    _     _
11    the       _     _     _     _     12    det     _     _
12    reunion   _     _     _     _     10    pobj    _     _
13    were      _     _     _     _     14    cop     _     _
14    dead      _     _     _     _     0     root    _     _
15    .         _     _     _     _     14    punct   _     _
```

(10) Natural disasters – storms, flooding, hurricanes – occur infrequently but cause devastation that strains resources to breaking point.

Roddy (+checked by Jennifer and Irene):

1	natural	_	_	_	_	2	amod	_	_
2	disasters	_	_	_	_	10	nsubj	_	_
3	-	_	_	_	_	4	punct	_	_
4	storms	_	_	_	_	2	appos	_	_
5	,	_	_	_	_	4	punct	_	_
6	flooding	_	_	_	_	4	conj	_	_
7	,	_	_	_	_	4	punct	_	_
8	hurricanes	_	_	_	_	4	conj	_	_
9	-	_	_	_	_	4	punct	_	_
10	occur	_	_	_	_	0	root	_	_
11	infrequently	_	_	_	_	10	advmod	_	_
12	but	_	_	_	_	10	cc	_	_
13	cause	_	_	_	_	10	conj	_	_
14	devastation	_	_	_	_	13	dobj	_	_
15	that	_	_	_	_	16	nsubj	_	_
16	strains	_	_	_	_	14	rcmod	_	_
17	resources	_	_	_	_	16	dobj	_	_
18	to	_	_	_	_	16	prep	_	_
19	breaking	_	_	_	_	20	amod	_	_
20	point	_	_	_	_	18	pobj	_	_
21	.	_	_	_	_	10	punct	_	_

(11) Letters delivered on time by old-fashioned means are increasingly rare, so it is as well that that is not the only option available.

Agreed by Inès and Jennifer (+Lucas, Daan) - discussion of 'so' here:

1	Letters	_	_	_	_	10	nsubj	_	_
2	delivered	_	_	_	_	1	vmod	_	_
3	on	_	_	_	_	2	prep	_	_
4	time	_	_	_	_	3	pobj	_	_
5	by	_	_	_	_	2	prep	_	_
6	old-fashioned	_	_	_	_	7	amod	_	_
7	means	_	_	_	_	5	pobj	_	_
8	are	_	_	_	_	10	cop	_	_
9	increasingly	_	_	_	_	10	advmod	_	_
10	rare	_	_	_	_	0	root	_	_
11	,	_	_	_	_	10	punct	_	_
12	so	_	_	_	_	10	cc	_	_
13	it	_	_	_	_	14	nsubj	_	_
14	is	_	_	_	_	16	cop	_	_
15	as	_	_	_	_	16	advmod	_	_
16	well	_	_	_	_	10	conj	_	_
17	that	_	_	_	_	23	mark	_	_
18	that	_	_	_	_	19	nsubj	_	_
19	is	_	_	_	_	23	cop	_	_
20	not	_	_	_	_	22	neg	_	_
21	the	_	_	_	_	23	det	_	_
22	only	_	_	_	_	23	amod	_	_
23	option	_	_	_	_	16	ccomp	_	_
24	available	_	_	_	_	23	amod	_	_
25	.	_	_	_	_	10	punct	_	_

(12) It won’t rain but there might be snow on high ground if the temperature stays about the same over the next 24 hours.

Jennifer + Lucas (+ Rowan)
Click here for discussion on “stays the same” - where did it go? :(
Discussion of ‘be’ being cop `here` and `here`

```
1     It          _     _     _     _     4     nsubj _     _
2     wo          _     _     _     _     4     aux   _     _
3     n’t         _     _     _     _     4     neg   _     _
4     rain        _     _     _     _     0     root  _     _
5     but         _     _     _     _     4     cc    _     _
6     there       _     _     _     _     8     expl  _     _
7     might       _     _     _     _     8     aux   _     _
8     be          _     _     _     _     4     conj  _     _
9     snow        _     _     _     _     8     nsubj _     _
10    on          _     _     _     _     9     prep  _     _
11    high        _     _     _     _     12    amod  _     _
12    ground            _     _     _     _     10    pobj  _     _
13    if          _     _     _     _     16    mark  _     _
14    the         _     _     _     _     15    det   _     _
15    temperature       _     _     _     _     16    nsubj _     _
16    stays       _     _     _     _     8     advcl _     _
17    about       _     _     _     _     19    advmod      _     _
18    the         _     _     _     _     19    det   _     _
19    same        _     _     _     _     16    acomp _     _
20    over        _     _     _     _     19    prep  _     _
21    the         _     _     _     _     24    det   _     _
22    next        _     _     _     _     24    amod  _     _
23    24          _     _     _     _     24    num   _     _
24    hours       _     _     _     _     20    pobj  _     _
25    .           _     _     _     _     4     punct _     _
```

(13) The long and lonely road to redemption begins with self-reflection: the need to delve inwards to deconstruct layers of psychological obfuscation.

Jennifer (+ Rowan + Yifan + Lucas): - discussion about appos + vmod discussion o and o

1	The	_	_	_	_	5	det	_	_	
2	long	_	_	_	_	5	amod	_	_	
3	and	_	_	_	_	2	cc	_	_	
4	lonely	_	_	_	_	2	conj	_	_	
5	road	_	_	_	_	8	nsubj	_	_	
6	to	_	_	_	_	5	prep	_	_	
7	redemption	_	_	_	_	6	pobj	_	_	
8	begins	_	_	_	_	0	root	_	_	
9	with	_	_	_	_	8	prep	_	_	
10	self-reflection	_	_	_	_	9	pobj	_	_	
11	:	_	_	_	_	8	punct	_	_	
12	the	_	_	_	_	13	det	_	_	
13	need	_	_	_	_	10	appos	_	_	
14	to	_	_	_	_	15	aux	_	_	
15	delve	_	_	_	_	13	vmod	_	_	
16	inwards	_	_	_	_	15	advmod	_	_	
17	to	_	_	_	_	18	aux	_	_	
18	deconstruct	_	_	_	_	15	vmod	_	_	
19	layers	_	_	_	_	18	dobj	_	_	
20	of	_	_	_	_	19	prep	_	_	
21	psychological	_	_	_	_	22	amod	_	_	
22	obfuscation	_	_	_	_	20	pobj	_	_	
23	.	_	_	_	_	8	punct	_	_	

(14) My wildest dream is to build a POS tagger which processes 10K words per second and uses only 1MB of RAM, but it may prove too hard.

Rowan (+ Jennifer + Kuan-Chi):

1	My	_	_	_	_	3	poss	_	_
2	wildest	_	_	_	_	3	amod	_	_
3	dream	_	_	_	_	4	nsubj	_	_
4	is	_	_	_	_	0	root	_	_
5	to	_	_	_	_	6	aux	_	_
6	build	_	_	_	_	4	xcomp	_	_
7	a	_	_	_	_	9	det	_	_
8	POS	_	_	_	_	9	nn	_	_
9	tagger	_	_	_	_	6	dobj	_	_
10	which	_	_	_	_	11	nsubj	_	_
11	processes	_	_	_	_	9	rcmod	_	_
12	10K	_	_	_	_	13	num	_	_
13	words	_	_	_	_	11	dobj	_	_
14	per	_	_	_	_	13	prep	_	_
15	second	_	_	_	_	14	pobj	_	_
16	and	_	_	_	_	11	cc	_	_
17	uses	_	_	_	_	11	conj	_	_
18	only	_	_	_	_	19	quantmod	_	_
19	1	_	_	_	_	20	num	_	_
20	MB	_	_	_	_	17	dobj	_	_
21	of	_	_	_	_	20	prep	_	_
22	RAM	_	_	_	_	21	pobj	_	_
23	,	_	_	_	_	4	punct	_	_
24	but	_	_	_	_	4	cc	_	_
25	it	_	_	_	_	27	nsubj	_	_
26	may	_	_	_	_	27	aux	_	_
27	prove	_	_	_	_	4	conj	_	_
28	too	_	_	_	_	29	advmod	_	_
29	hard	_	_	_	_	27	acomp	_	_
30	.	_	_	_	_	4	punct	_	_

(15) English also has many words of more or less unique function, including interjections (oh, ah), negatives (no, not), politeness markers (please, thank you), and the existential 'there' (there are horses but not unicorns) among others.

Agreed by Roddy + Lucas (+checked by Jennifer (+ okayed by rowan)):

1	English	_	_	_	_	3	nsubj	_	_		
2	also	_	_	_	_	3	advmod	_	_		
3	has	_	_	_	_	0	root	_	_		
4	many	_	_	_	_	5	amod	_	_		
5	words	_	_	_	_	3	dobj	_	_		
6	of	_	_	_	_	5	prep	_	_		
7	more	_	_	_	_	10	advmod	_	_		
8	or	_	_	_	_	7	cc	_	_	_	_
9	less	_	_	_	_	7	conj	_	_		
10	unique	_	_	_	_	11	amod	_	_		
11	function	_	_	_	_	6	pobj	_	_		
12	,	_	_	_	_	3	punct	_	_		
13	including	_	_	_	_	5	vmod	_	_		
14	interjections	_	_	_	_	13	dobj	_	_		
15	(	_	_	_	_	16	punct	_	_		
16	oh	_	_	_	_	14	appos	_	_		
17	,	_	_	_	_	16	punct	_	_		
18	ah	_	_	_	_	16	conj	_	_		
19	)	_	_	_	_	16	punct	_	_		
20	,	_	_	_	_	14	punct	_	_		
21	negatives	_	_	_	_	14	conj	_	_		
22	(	_	_	_	_	23	punct	_	_		
23	no	_	_	_	_	21	appos	_	_		
24	,	_	_	_	_	23	punct	_	_		
25	not	_	_	_	_	23	conj	_	_		
26	)	_	_	_	_	23	punct	_	_		
27	,	_	_	_	_	14	punct	_	_		
28	politeness	_	_	_	_	29	nn	_	_		
29	markers	_	_	_	_	14	conj	_	_		
30	(	_	_	_	_	31	punct	_	_		
31	please	_	_	_	_	29	appos	_	_		
32	,	_	_	_	_	31	punct	_	_		
33	thank	_	_	_	_	31	conj	_	_		
34	you	_	_	_	_	33	dobj	_	_		
35	)	_	_	_	_	31	punct	_	_		
36	,	_	_	_	_	14	punct	_	_		
37	and	_	_	_	_	14	cc	_	_		
38	the	_	_	_	_	41	det	_	_		
39	existential	_	_	_	_	41	amod	_	_		
40	'	_	_	_	_	41	punct	_	_		
41	there	_	_	_	_	14	conj	_	_		
42	'	_	_	_	_	41	punct	_	_		
43	(	_	_	_	_	45	punct	_	_		
44	there	_	_	_	_	45	expl	_	_		
45	are	_	_	_	_	41	dep	_	_		
46	horses	_	_	_	_	45	nsubj	_	_		

47	but	_	_	_	_	46	cc	_	_
48	not	_	_	_	_	49	neg	_	_
49	unicorns	_	_	_	_	46	conj	_	_
50	)	_	_	_	_	45	punct	_	_
51	among	_	_	_	_	14	prep	_	_
52	others	_	_	_	_	51	pobj	_	_
53	.	_	_	_	_	3	punct	_	_

(16) Making these decisions requires sophisticated knowledge of syntax; tagging manuals (Santorini, 1990) give various heuristics that can help human coders make these decisions and that can also provide useful features for automatic taggers.

Agreed by Stefan and Yifan (+checked by Jennifer):

```
1    Making      _    _    _    _    4    csubj       _    _
2    these       _    _    _    _    3    det         _    _
3    decisions   _    _    _    _    1    dobj        _    _
4    requires    _    _    _    _    0    root        _    _
5    sophisticated    _    _    _    _    6    amod        _    _
6    knowledge   _    _    _    _    4    dobj        _    _
7    of          _    _    _    _    6    prep        _    _
8    syntax      _    _    _    _    7    pobj        _    _
9    ;           _    _    _    _    4    punct       _    _
10   tagging     _    _    _    _    11   amod        _    _
11   manuals     _    _    _    _    17   nsubj       _    _
12   (           _    _    _    _    11   punct       _    _
13   Santorini   _    _    _    _    11   appos       _    _
14   ,           _    _    _    _    13   punct       _    _
15   1990        _    _    _    _    13   dep         _    _
16   )           _    _    _    _    11   punct       _    _
17   give        _    _    _    _    4    parataxis   _    _
18   various     _    _    _    _    19   amod        _    _
19   heuristics  _    _    _    _    17   dobj        _    _
20   that        _    _    _    _    22   nsubj       _    _
21   can         _    _    _    _    22   aux         _    _
22   help        _    _    _    _    19   rcmod       _    _
23   human       _    _    _    _    24   amod        _    _
24   coders      _    _    _    _    22   dobj        _    _
25   make        _    _    _    _    22   xcomp       _    _
26   these       _    _    _    _    27   det         _    _
27   decisions   _    _    _    _    25   dobj        _    _
28   and         _    _    _    _    22   cc          _    _
29   that        _    _    _    _    32   nsubj       _    _
30   can         _    _    _    _    32   aux         _    _
31   also        _    _    _    _    32   advmod      _    _
32   provide     _    _    _    _    22   conj        _    _
33   useful      _    _    _    _    34   amod        _    _
34   features    _    _    _    _    32   dobj        _    _
35   for         _    _    _    _    34   prep        _    _
36   automatic   _    _    _    _    37   amod        _    _
37   taggers     _    _    _    _    35   pobj        _    _
38   .           _    _    _    _    4    punct       _    _
```

(17) The Penn Treebank tagset was culled from the original 87-tag tagset for the Brown Corpus. For example the original Brown and C5 tagsets include a separate tag for each of the different forms of the verbs *do* (e.g. C5 tag VDD for did and VDG tag for doing), *be* and *have*.

Agreed by Irene and Yifan (+checked by Jennifer):

```
1     The       _     _     _     _     4     det       _     _
2     Penn      _     _     _     _     4     nn        _     _
3     Treebank  _     _     _     _     4     nn        _     _
4     tagset    _     _     _     _     6     nsubjpass _     _
5     was       _     _     _     _     6     auxpass   _     _
6     culled    _     _     _     _     0     root      _     _
7     from      _     _     _     _     6     prt       _     _
8     the       _     _     _     _     11    det       _     _
9     original  _     _     _     _     11    amod      _     _
10    87-tag    _     _     _     _     11    amod      _     _
11    tagset    _     _     _     _     7     pobj      _     _
12    for       _     _     _     _     11    prep      _     _
13    the       _     _     _     _     15    det       _     _
14    Brown     _     _     _     _     15    nn        _     _
15    Corpus    _     _     _     _     12    pobj      _     _
16    .         _     _     _     _     6     punct     _     _
--------
1     For       _     _     _     _     9     prep      _     _
2     example   _     _     _     _     1     pobj      _     _
3     the       _     _     _     _     8     det       _     _
4     original  _     _     _     _     8     amod      _     _
5     Brown     _     _     _     _     8     nn        _     _
6     and       _     _     _     _     5     cc        _     _
7     C5        _     _     _     _     5     conj      _     _
8     tagsets   _     _     _     _     9     nsubj     _     _
9     include   _     _     _     _     0     root      _     _
10    a         _     _     _     _     12    det       _     _
11    separate  _     _     _     _     12    amod      _     _
12    tag       _     _     _     _     9     dobj      _     _
13    for       _     _     _     _     12    prep      _     _
14    each      _     _     _     _     13    pobj      _     _
15    of        _     _     _     _     14    prep      _     _
16    the       _     _     _     _     18    det       _     _
17    different _     _     _     _     18    amod      _     _
18    forms     _     _     _     _     13    pobj      _     _
19    of        _     _     _     _     18    prep      _     _
20    the       _     _     _     _     21    det       _     _
21    verbs     _     _     _     _     22    nn        _     _
22    do        _     _     _     _     19    pobj      _     _
```

```
23    (      _     _     _     _     22    punct _     _
24    e.g.   _     _     _     _     27    nn    _     _
25    C5     _     _     _     _     27    nn    _     _
26    tag    _     _     _     _     27    nn    _     _
27    VDD    _     _     _     _     22    appos _     _
28    for    _     _     _     _     27    prep  _     _
29    did    _     _     _     _     28    pobj  _     _
30    and    _     _     _     _     27    cc    _     _
31    VDG    _     _     _     _     32    nn    _     _
32    tag    _     _     _     _     27    conj  _     _
33    for    _     _     _     _     32    prep  _     _
34    doing  _     _     _     _     33    pobj  _     _
35    )      _     _     _     _     22    punct _     _
36    ,      _     _     _     _     22    punct _     _
37    be     _     _     _     _     22    conj  _     _
38    and    _     _     _     _     22    cc    _     _
39    have   _     _     _     _     22    conj  _     _
40    .      _     _     _     _     9     punct _     _
```

(18) The slightly simplified version of the Viterbi algorithm that we present takes as input a single HMM and a sequence of observed words O = (o1,o2,...oT) and returns the most probable state/tag sequence Q = (q1,q2,qT) together with its probability.

Agreed by Irene, Inès, Kuan-Chi and Jennifer:

```
1     The        _     _     _     _     4     det     _     _
2     slightly   _     _     _     _     3     advmod  _     _
3     simplified _     _     _     _     4     amod    _     _
4     version    _     _     _     _     12    nsubj   _     _
5     of         _     _     _     _     4     prep    _     _
6     the        _     _     _     _     8     det     _     _
7     Viterbi    _     _     _     _     8     nn      _     _
8     algorithm  _     _     _     _     5     pobj    _     _
9     that       _     _     _     _     11    mark    _     _
10    we         _     _     _     _     11    nsubj   _     _
11    present    _     _     _     _     4     rcmod   _     _
12    takes      _     _     _     _     0     root    _     _
13    as         _     _     _     _     12    prep    _     _
14    input      _     _     _     _     13    pobj    _     _
15    a          _     _     _     _     17    det     _     _
16    single     _     _     _     _     17    amod    _     _
17    HMM        _     _     _     _     12    dobj    _     _
18    and        _     _     _     _     17    cc      _     _
```

19	a	_	_	_	_	20	det	_	_		
20	sequence	_	_	_	_	17	conj	_	_		
21	of	_	_	_	_	20	prep	_	_		
22	observed	_	_	_	_	23	amod	_	_		
23	words	_	_	_	_	21	pobj	_	_		
24	O=(o1,..,oT)	_	_	_	_	23	appos	_	_		
25	and	_	_	_	_	12	cc	_	_		
26	returns	_	_	_	_	12	conj	_	_		
27	the	_	_	_	_	33	det	_	_		
28	most	_	_	_	_	29	advmod	_	_		
29	probable	_	_	_	_	33	amod	_	_		
30	state	_	_	_	_	33	nn	_	_		
31	/	_	_	_	_	30	cc	_	_		
32	tag	_	_	_	_	30	conj	_	_		
33	sequence	_	_	_	_	26	dobj	_	_		
34	*Q=(q1,..qT)*	_	_	_	_	*33*	*appos*	_	_		
35	together	_	_	_	_	36	advmod	_	_		
36	with	_	_	_	_	33	prep	_	_		
37	its	_	_	_	_	38	poss	_	_		
38	probability	_	_	_	_	36	pobj	_	_		
39	.	_	_	_	_	12	punct	_	_		

(19) Thus the EM-trained “pure HMM” tagger is probably best suited to cases where no training data is available, for example, when tagging languages for which no data was previously hand-tagged.
(Inès+checked by Jennifer and Irene)

1	Thus	_	_	_	_	12	advmod	_	_
2	the	_	_	_	_	8	det	_	_
3	EM-trained	_	_	_	_	8	amod	_	_
4	“	_	_	_	_	6	punct	_	_
5	pure	_	_	_	_	6	amod	_	_
6	HMM	_	_	_	_	8	nn	_	_
7	”	_	_	_	_	6	punct	_	_
8	tagger	_	_	_	_	12	nsubj	_	_
9	is	_	_	_	_	12	aux	_	_
10	probably	_	_	_	_	12	advmod	_	_
11	best	_	_	_	_	12	amod	_	_
12	suited	_	_	_	_	0	root	_	_
13	to	_	_	_	_	12	prep	_	_
14	cases	_	_	_	_	13	pobj	_	_
15	where	_	_	_	_	20	advmod	_	_
16	no	_	_	_	_	18	neg	_	_
17	training	_	_	_	_	18	amod	_	_
18	data	_	_	_	_	20	nsubj	_	_
19	is	_	_	_	_	20	cop	_	_
20	available	_	_	_	_	14	advcl	_	_
21	,	_	_	_	_	14	punct	_	_
22	for	_	_	_	_	14	prep	_	_
23	example	_	_	_	_	22	pobj	_	_
24	,	_	_	_	_	14	punct	_	_
25	when	_	_	_	_	26	advmod	_	_
26	tagging	_	_	_	_	12	advcl	_	_
27	languages	_	_	_	_	26	dobj	_	_
28	for	_	_	_	_	34	prep	_	_
29	which	_	_	_	_	28	dobj	_	_
30	no	_	_	_	_	31	neg	_	_
31	data	_	_	_	_	34	nsubjpass	_	_
32	was	_	_	_	_	34	auxpass	_	_
33	previously	_	_	_	_	34	advmod	_	_
34	hand-tagged	_	_	_	_	27	rcmod	_	_
35	.	_	_	_	_	12	punct	_	_

(20) Coming home from very lonely places, all of us go a little mad: whether from great personal success, or just an all-night drive, we are the sole survivors of a world no one else has ever seen.

Jennifer (+checked by Kuan-Chi + Rowan):

1	Coming	_	_	_	_	11	vmod	_	_	
2	home	_	_	_	_	1	xcomp	_	_	
3	from	_	_	_	_	2	prep	_	_	
4	very	_	_	_	_	5	advmod	_	_	
5	lonely	_	_	_	_	6	amod	_	_	
6	places	_	_	_	_	3	pobj	_	_	
7	,	_	_	_	_	12	punct	_	_	
8	all	_	_	_	_	11	nsubj	_	_	
9	of	_	_	_	_	8	prep	_	_	
10	us	_	_	_	_	9	pobj	_	_	
11	go	_	_	_	_	0	root	_	_	
12	a	_	_	_	_	13	det	_	_	
13	little	_	_	_	_	14	npadvmod	_	_	
14	mad	_	_	_	_	11	xcomp	_	_	
15	:	_	_	_	_	11	punct	_	_	
16	whether	_	_	_	_	17	mark	_	_	
17	from	_	_	_	_	14	prep	_	_	
18	great	_	_	_	_	20	amod	_	_	
19	personal	_	_	_	_	20	amod	_	_	
20	success	_	_	_	_	17	pobj	_	_	
21	,	_	_	_	_	17	punct	_	_	
22	or	_	_	_	_	20	cc	_	_	
23	just	_	_	_	_	26	advmod	_	_	
24	an	_	_	_	_	26	det	_	_	
25	all-night	_	_	_	_	26	amod	_	_	
26	drive	_	_	_	_	20	conj	_	_	
27	,	_	_	_	_	26	punct	_	_	
28	we	_	_	_	_	29	nsubj	_	_	
29	are	_	_	_	_	32	cop	_	_	
30	the	_	_	_	_	32	det	_	_	
31	sole	_	_	_	_	32	amod	_	_	
32	survivors	_	_	_	_	11	parataxis	_	_	
33	of	_	_	_	_	32	prep	_	_	
34	a	_	_	_	_	35	det	_	_	
35	world	_	_	_	_	33	pobj	_	_	
36	no	_	_	_	_	37	neg	_	_	
37	one	_	_	_	_	41	nsubj	_	_	
38	else	_	_	_	_	37	advmod	_	_	
39	has	_	_	_	_	41	aux	_	_	
40	ever	_	_	_	_	41	advmod	_	_	

```
41    seen _     _     _     _     35    rcmod _     _
42    .    _     _     _     _     11    punct _     _
```

(21) Skill without imagination is craftsmanship and gives us many useful objects such as wickerwork picnic baskets. Imagination without skill gives us modern art.

Rami (+ checked by Jennifer and Irene + Rowan):

1	Skill	_	_	_	_	5	nsubj	_	_	
2	without	_	_	_	_	1	prep	_	_	
3	imagination	_	_	_	_	2	pobj	_	_	
4	is	_	_	_	_	5	cop	_	_	
5	craftsmanship	_	_	_	_	0	root	_	_	
6	and	_	_	_	_	5	cc	_	_	
7	gives	_	_	_	_	5	conj	_	_	
8	us	_	_	_	_	7	iobj	_	_	
9	many	_	_	_	_	11	amod	_	_	
10	useful	_	_	_	_	11	amod	_	_	
11	objects	_	_	_	_	7	dobj	_	_	
12	such	_	_	_	_	13	mwe	_	_	
13	as	_	_	_	_	11	prep	_	_	
14	wickerwork	_	_	_	_	16	nn	_	_	
15	picnic	_	_	_	_	16	nn	_	_	
16	baskets	_	_	_	_	13	pobj	_	_	
17	.	_	_	_	_	5	punct	_	_	
18	Imagination	_	_	_	_	21	nsubj	_	_	
19	without	_	_	_	_	18	prep	_	_	
20	skill	_	_	_	_	19	pobj	_	_	
21	gives	_	_	_	_	0	root	_	_	
22	us	_	_	_	_	21	iobj	_	_	
23	modern	_	_	_	_	24	amod	_	_	
24	art	_	_	_	_	21	dobj	_	_	
25	.	_	_	_	_	21	punct	_	_	

(22) An MoD spokesman said: “Surveys of Astute have now been completed and she will proceed to Faslane under her own power. She is being escorted by tugs and HMS Shoreham.”

Lucas (+checked by Jennifer + Rowan) discussion about 2nd parataxis here . punct here.ccomp for quotation here

1	An	_	_	_	_	3	det	_	_	
2	MoD	_	_	_	_	3	nn	_	_	
3	spokesman	_	_	_	_	4	nsubj	_	_	
4	said	_	_	_	_	0	root	_	_	
5	:	_	_	_	_	4	punct	_	_	
6	``	_	_	_	_	4	punct	_	_	
7	Surveys	_	_	_	_	13	nsubjpass	_	_	
8	of	_	_	_	_	7	prep	_	_	
9	Astute	_	_	_	_	8	pobj	_	_	
10	have	_	_	_	_	13	aux	_	_	
11	now	_	_	_	_	13	advmod	_	_	
12	been	_	_	_	_	13	auxpass	_	_	
13	completed	_	_	_	_	4	ccomp	_	_	
14	and	_	_	_	_	13	cc	_	_	
15	she	_	_	_	_	17	nsubj	_	_	
16	will	_	_	_	_	17	aux	_	_	
17	proceed	_	_	_	_	13	conj	_	_	
18	to	_	_	_	_	17	prep	_	_	
19	Faslane	_	_	_	_	18	pobj	_	_	
20	under	_	_	_	_	17	prep	_	_	
21	her	_	_	_	_	23	poss	_	_	
22	own	_	_	_	_	23	amod	_	_	
23	power	_	_	_	_	20	pobj	_	_	
24	.	_	_	_	_	13	punct	_	_	
25	She	_	_	_	_	28	nsubjpass	_	_	
26	is	_	_	_	_	28	aux	_	_	
27	being	_	_	_	_	28	auxpass	_	_	
28	escorted	_	_	_	_	13	parataxis	_	_	
29	by	_	_	_	_	28	prep	_	_	
30	tugs	_	_	_	_	29	pobj	_	_	
31	and	_	_	_	_	30	cc	_	_	
32	HMS	_	_	_	_	33	nn	_	_	
33	Shoreham	_	_	_	_	30	conj	_	_	
34	.	_	_	_	_	28	punct	_	_	
35	’’	_	_	_	_	4	punct	_	_	

(23) But far fewer people fully understand how the Media Lab operates, fits into MIT, and encourages such a creative environment; about half of the anniversary celebration's program focused on simply defining what the Media Lab is.

Rowan (+ checked by Jennifer,Yifan):

1	But	_	_	_	_	6	cc	_	_
2	far	_	_	_	_	3	advmod	_	_
3	fewer	_	_	_	_	4	amod	_	_
4	people	_	_	_	_	6	nsubj	_	_
5	fully	_	_	_	_	6	advmod	_	_
6	understand	_	_	_	_	0	root	_	_
7	how	_	_	_	_	11	advmod	_	_
8	the	_	_	_	_	10	det	_	_
9	Media	_	_	_	_	10	nn	_	_
10	Lab	_	_	_	_	11	nsubj	_	_
11	operates	_	_	_	_	6	ccomp	_	_
12	,	_	_	_	_	10	punct	_	_
13	fits	_	_	_	_	11	conj	_	_
14	into	_	_	_	_	13	prep	_	_
15	MIT	_	_	_	_	14	pobj	_	_
16	,	_	_	_	_	10	punct	_	_
17	and	_	_	_	_	11	cc	_	_
18	encourages	_	_	_	_	11	conj	_	_
19	such	_	_	_	_	22	predet	_	_
20	a	_	_	_	_	22	det	_	_
21	creative	_	_	_	_	22	amod	_	_
22	environment	_	_	_	_	18	dobj	_	_
23	;	_	_	_	_	6	punct	_	_
24	about	_	_	_	_	25	quantmod	_	_
25	half	_	_	_	_	31	nsubj	_	_
26	of	_	_	_	_	25	prep	_	_
27	the	_	_	_	_	29	det	_	_
28	anniversary	_	_	_	_	29	nn	_	_
29	celebration	_	_	_	_	31	poss	_	_
30	's	_	_	_	_	29	possessive	_	_
31	program	_	_	_	_	25	pobj	_	_
32	focused	_	_	_	_	6	parataxis	_	_
33	on	_	_	_	_	32	prep	_	_
34	simply	_	_	_	_	35	advmod	_	_
35	defining	_	_	_	_	33	pcomp	_	_
36	what	_	_	_	_	40	dobj	_	_
37	the	_	_	_	_	39	det	_	_
38	Media	_	_	_	_	39	nn	_	_
39	Lab	_	_	_	_	40	nsubj	_	_
40	is	_	_	_	_	35	ccomp	_	_
41	.	_	_	_	_	6	punct	_	_

(24) Instead of constantly worrying about funding, the faculty and students can focus on their project, with the exception of sponsors' weeks, when they have to convince companies to start or continue their support.

Roddy:

```
1    instead     _    _    _    _    2    mwe    _    _
2    of    _     _    _    _    13   prep   _    _
3    constantly  _    _    _    _    4    advmod      _    _
4    worrying    _    _    _    _    2    pcomp  _    _
5    about  _    _    _    _    4    prep   _    _
6    funding     _    _    _    _    5    pobj   _    _
7    ,     _     _    _    _    4    punct  _    _
8    the   _     _    _    _    9    det    _    _
9    faculty     _    _    _    _    13   nsubj  _    _
10   and   _     _    _    _    9    cc     _    _
11   students    _    _    _    _    9    conj   _    _
12   can   _     _    _    _    13   aux    _    _
13   focus  _    _    _    _    0    root   _    _
14   on    _     _    _    _    13   prep   _    _
15   their  _    _    _    _    16   poss   _    _
16   project     _    _    _    _    14   pobj   _    _
17   ,     _     _    _    _    13   punct  _    _
18   with  _     _    _    _    13   prep   _    _
19   the   _     _    _    _    20   det    _    _
20   exception   _    _    _    _    18   pobj   _    _
21   of    _     _    _    _    20   prep   _    _
22   sponsors    _    _    _    _    24   poss   _    _
23   '     _     _    _    _    22   possessive  _    _
24   weeks  _    _    _    _    21   pobj   _    _
25   ,     _     _    _    _    13   punct  _    _
26   when  _     _    _    _    28   mark   _    _
27   they  _     _    _    _    28   nsubj  _    _
28   have  _     _    _    _    24   rcmod  _    _
29   to    _     _    _    _    30   aux    _    _
30   convince    _    _    _    _    28   xcomp  _    _
31   companies   _    _    _    _    30   dobj   _    _
32   to    _     _    _    _    33        aux    _
33   start  _    _    _    _    30   xcomp  _    _
34   or    _     _    _    _    33   cc     _    _
35   continue    _    _    _    _    33   conj   _    _
36   their  _    _    _    _    37   poss   _    _
37   support     _    _    _    _    33   dobj   _    _
38   .     _     _    _    _    13   punct  _    _
```

(25) The doctors are warning that the NHS cannot make the £20bn of savings by 2014 that ministers expect, while simultaneously undertaking a huge reorganisation that will see England's 152 primary care trusts (PCTs) abolished and consortiums of GPs assume responsibility for the commissioning of services for patients.

Agreed by Kuan-Chi, Stefan, and Yifan:

1	The	_	_	_	_	2	det	_	_	
2	doctors	_	_	_	_	4	nsubj	_	_	
3	are	_	_	_	_	4	aux	_	_	
4	warning	_	_	_	_	0	root	_	_	
5	that	_	_	_	_	10	mark	_	_	
6	the	_	_	_	_	7	det	_	_	
7	NHS	_	_	_	_	10	nsubj	_	_	
8	can	_	_	_	_	10	aux	_	_	
9	not	_	_	_	_	10	neg	_	_	
10	make	_	_	_	_	4	ccomp	_	_	
11	the	_	_	_	_	13	det	_	_	
12	£	_	_	_	_	10	dobj	_	_	
13	20bn	_	_	_	_	12	num	_	_	
14	of	_	_	_	_	13	prep	_	_	
15	savings	_	_	_	_	14	pobj	_	_	
16	by	_	_	_	_	10	prep	_	_	
17	2014	_	_	_	_	16	pobj	_	_	
18	that	_	_	_	_	20	dobj	_	_	
19	ministers	_	_	_	_	20	nsubj	_	_	
20	expect	_	_	_	_	12	rcmod	_	_	
21	,	_	_	_	_	4	punct	_	_	
22	while	_	_	_	_	10	mark	_	_	
23	simultaneously	_	_	_	_	24	advmod	_	_	
24	undertaking	_	_	_	_	22	xcomp	_	_	
25	a	_	_	_	_	27	det	_	_	
26	huge	_	_	_	_	27	amod	_	_	
27	reorganisation	_	_	_	_	24	dobj	_	_	
28	that	_	_	_	_	30	nsubj	_	_	
29	will	_	_	_	_	30	aux	_	_	
30	see	_	_	_	_	27	rcmod	_	_	
31	England	_	_	_	_	36	poss	_	_	
32	's	_	_	_	_	31	possessive	_	_	
33	152	_	_	_	_	36	num	_	_	
34	primary	_	_	_	_	35	amod	_	_	
35	care	_	_	_	_	36	nn	_	_	
36	trusts	_	_	_	_	30	dobj	_	_	
37	(	_	_	_	_	38	punct	_	_	
38	PCTs	_	_	_	_	36	appos	_	_	
39	)	_	_	_	_	38	punct	_	_	

```
40  abolished  _  _  _  _  30  xcomp  _  _
41  and  _  _  _  _  40  cc  _  _
42  consortiums  _  _  _  _  45  nsubj  _  _
43  of  _  _  _  _  42  prep  _  _
44  GPs  _  _  _  _  43  pobj  _  _
45  assume  _  _  _  _  40  conj  _  _
46  responsibility  _  _  _  _  45  dobj  _  _
47  for  _  _  _  _  46  prep  _  _
48  the  _  _  _  _  49  det  _  _
49  commissioning  _  _  _  _  47  pobj  _  _
50  of  _  _  _  _  49  prep  _  _
51  services  _  _  _  _  50  pobj  _  _
52  for  _  _  _  _  51  prep  _  _
53  patients  _  _  _  _  52  pobj  _  _
54  .  _  _  _  _  4  punct  _  _
```