

Functions and Objectives of a Query Language for User- Activity Data

Michael Carl
CBS - CRITT

Overview

- Representation of User Activity Data:
 - TRANSLOG (key-logging)
 - GWM (eye-tracking)
 - Textual data
- Elements of a Query Language for UAD
 - Linking Product and Process Data
 - What can be retrieved
 - What would we like to know from UAD

Spielberg shows Beijing red card over Darfur

In a gesture sure to rattle the Chinese Government, Steven Spielberg pulled out of the Beijing Olympics to protest against China's backing for Sudan's policy in Darfur. His withdrawal comes in the wake of fighting flaring up again in Darfur and is set to embarrass China, which has sought to halt the negative fallout from having close ties to the Sudanese government. China, which has extensive investments in the Sudanese oil industry, maintains close links with the Government, which includes one minister charged with crimes against humanity by the International Criminal Court in The Hague. Although emphasizing that Khartoum bears the bulk of the responsibility for these ongoing atrocities, Spielberg maintains that the international community, and particularly China, should do more to end the suffering.

Spielberg giver Beijing det røde kort for Darfur

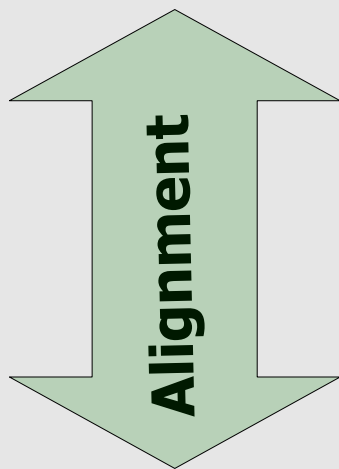
I en der helt sikkert ryster den kinesiske regering, trak Steven Spielberg sig fra de olympiske lege i Beijing for at protestere imod Kinas støtte af den sudanske...i Darfur. Hans tilbagetrækning kommer ikampene i Darfur

Product Data

Process Data

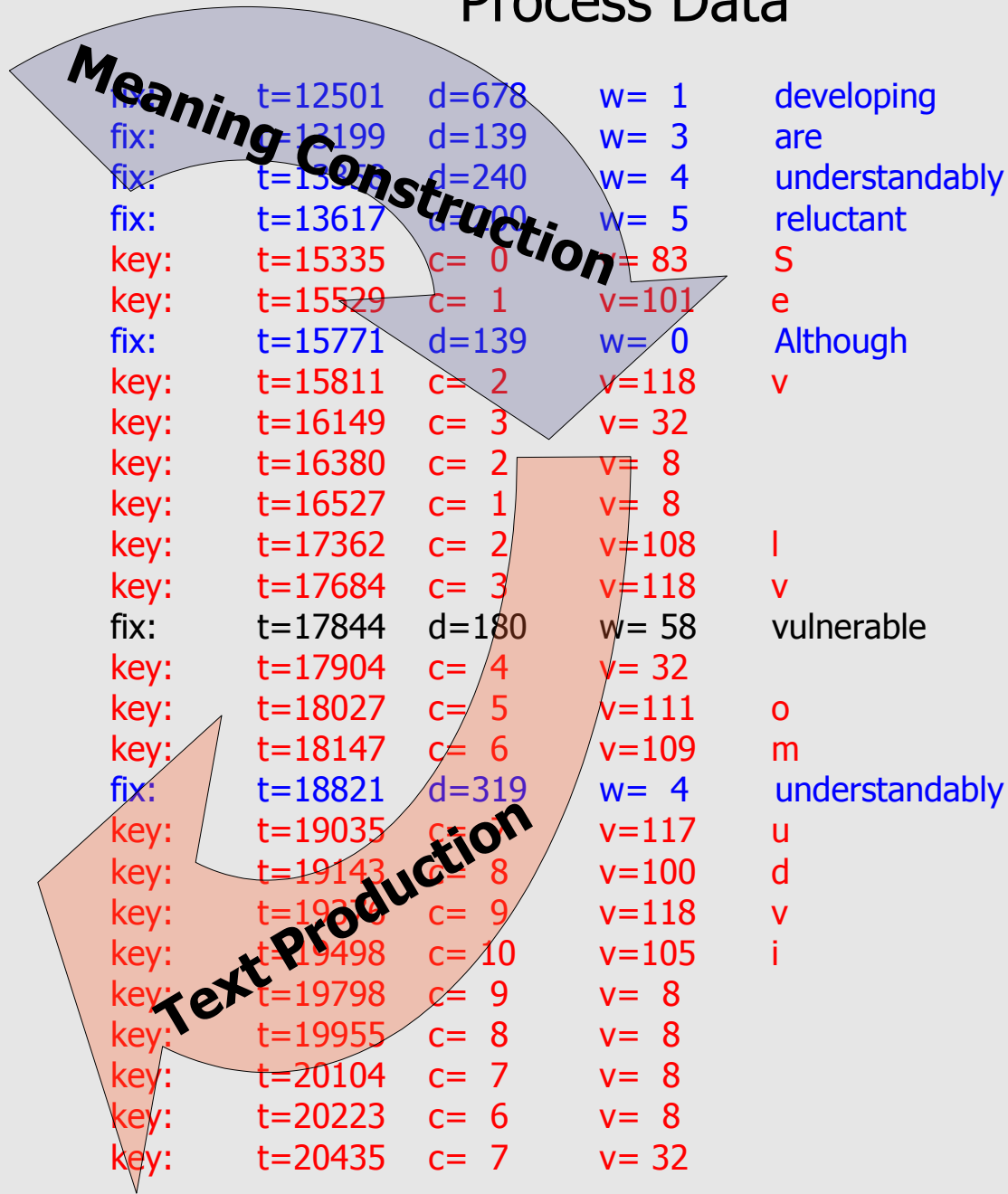
Source

Although developing countries are understandably reluctant to compromise ...



Selv om udviklingslandene forståeligt nok er modvillige mod at risikere deres chancer ...

Target



Structure of the UA Data

- Product Data (static):
 - spacial distribution of textual items (words, phrases)
 - Source Text
 - Target Text
 - Alignment
- Process Data (dynamic):
 - temporal distribution of tracking & logging data
 - Eye-gaze Samples
 - Fixations (and Fixation -> Word Mappings)
 - Keyboard Data

Representation UA Data

- Source text:

```
<word cur="9" pos="VVG" val="developing"/>
<word cur="20" pos="NN2" val="countries" />
```

- Target text:

```
<word cur="0" val="Selv" ... />
<word cur="9" val="udviklingslandene" ... />
```

- Alignment

```
<align source="9 20" target="9" ... />
```

- Eye-gaze samples points:

```
<eye time="140" lx="748.48" ly="1222.29" lp="3.85335" ... />
<eye time="160" lx="757.561" ly="1213.05" lp="3.78815" ... />
```

- Fixation to word mappings:

```
<fix time="12501" dur="678" cur="9" ... />
<fix time="13199" dur="139" cur="30" ... />
```

- Keyboard activities:

```
<key time="15346" val="83" cur="1" ... />
<key time="15540" val="101" cur="2" ... />
```



Objectives for a Query Language

- Discover structures (or patterns) in UAD:
 - Product data (monolingual, bilingual)
 - Process data (keystrokes, gaze data)
- Quantify dependencies in UAD:
 - unconditional/conditional
 - between process & product data
- Elaborate a Model of Translation Activities:
 - uncover translation difficulties
 - predict translation behavior

Functions of a Query Language

- Describe patterns in one modality:
 - Key, Fixation, Source, Target, Alignment
 - Retrieve all instances in the data that satisfy the pattern
- Link patterns from different modalities:
 - Via spacial order (cursor position):
 - from source to target: alignment
 - from target to keystrokes
 - Via temporal ordering
 - fixations & keyboard actions
- Pipeline process:
 - detect patterns in one modality
 - filter patterns through patterns of further modalities (conditional patterns)

Ex1: Keyboard Pattern

```
K>k[time.V.$T1],[time.V.$T2]  
!:if($T2-$T1<500) {PrintPattern('k');}
```

```
k:0   cur=1   time=15346   val=83 chr=S  
k:1   cur=2   time=15540   val=101 chr=e  
k:2   cur=3   time=15771   val=118 chr=v  
k:3   cur=4   time=16109   val=32  
k:4   cur=3   time=16340   val=8  
----- 835  
k:6   cur=3   time=17322   val=108 chr=l  
k:7   cur=4   time=17644   val=118 chr=v  
k:8   cur=5   time=17901   val=32  
k:9   cur=6   time=18024   val=111 chr=o
```

Ex2: Fixation Patterns

```
F>[time.V.$T1,cur.V.$C1],[time.V.$T2]  
K>&k[time.ge.$T1,time.le.$T2]  
S>src[cur.=.$C1]  
!:PrintPattern('src');
```

```
src:39 cur=277 str=extra  
src:33 cur=236 str=encourage  
src:30 cur=222 str=be  
src:28 cur=206 str=Incentives  
src:31 cur=225 str=offered  
src:33 cur=236 str=encourage  
src:34 cur=246 str=developing  
src:38 cur=273 str=the t  
src:43 cur=298 str=implement
```

```
src:47 cur=332 str=could  
src:48 cur=338 str=also  
src:46 cur=328 str=and  
src:28 cur=206 str=Incentives  
src:31 cur=225 str=offered  
src:47 cur=332 str=could  
src:49 cur=343 str=help  
src:49 cur=343 str=help  
src:50 cur=348 str=minimise  
src:47 cur=332 str=could  
src:50 cur=348 str=minimise
```

-

....

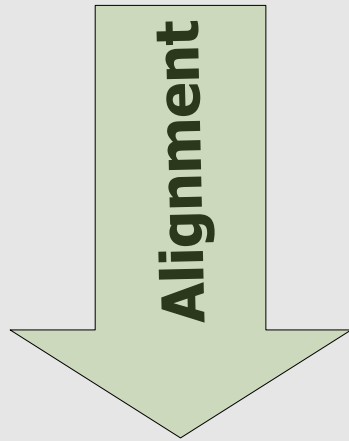
development. Incentives must be offered to encourage developing countries to go the extra green mile and implement clean technologies, and could also help minimise emissions from deforestation. Some of the most vulnerable countries

JUMP	Time	Pupil Dilat.	Duration	Cursor	Word#
0	t=120162	p=3.48574	d=139	c=277	n=39 extra
-41	t=120361	p=3.47241	d=120	c=236	n=33 encourage
-14	t=120541	p=3.50274	d=139	c=222	n=30 be
-16	t=120720	p=3.51766	d=179	c=206	n=28 Incentives
19	t=120919	p=3.35288	d=279	c=225	n=31 offered
11	t=121238	p=3.1657	d=220	c=236	n=33 encourage
10	t=121518	p=3.15666	d=199	c=246	n=34 developing
27	t=121757	p=3.15974	d=458	c=273	n=38 the
25	t=122315	p=3.15768	d=239	c=298	n=43 implement
10	t=122574	p=3.198	d=459	c=308	n=44 clean
24	t=123113	p=3.16391	d=239	c=332	n=47 could
6	t=123372	p=3.12157	d=279	c=338	n=48 also
-10	t=123671	p=3.1349	d=259	c=328	n=46 and
-122	t=124010	p=3.08586	d=239	c=206	n=28 Incentives
19	t=124309	p=3.11961	d=139	c=225	n=31 offered
107	t=124468	p=3.1269	d=220	c=332	n=47 could
11	t=124708	p=3.21573	d=318	c=343	n=49 help
0	t=126243	p=3.26034	d=239	c=343	n=49 help
5	t=126502	p=3.26961	d=399	c=348	n=50 minimise
-16	t=127898	p=3.27108	d=378	c=332	n=47 could
16	t=128296	p=3.27849	d=319	c=348	n=50 minimise

Ex3: Translation Production Patterns

Source

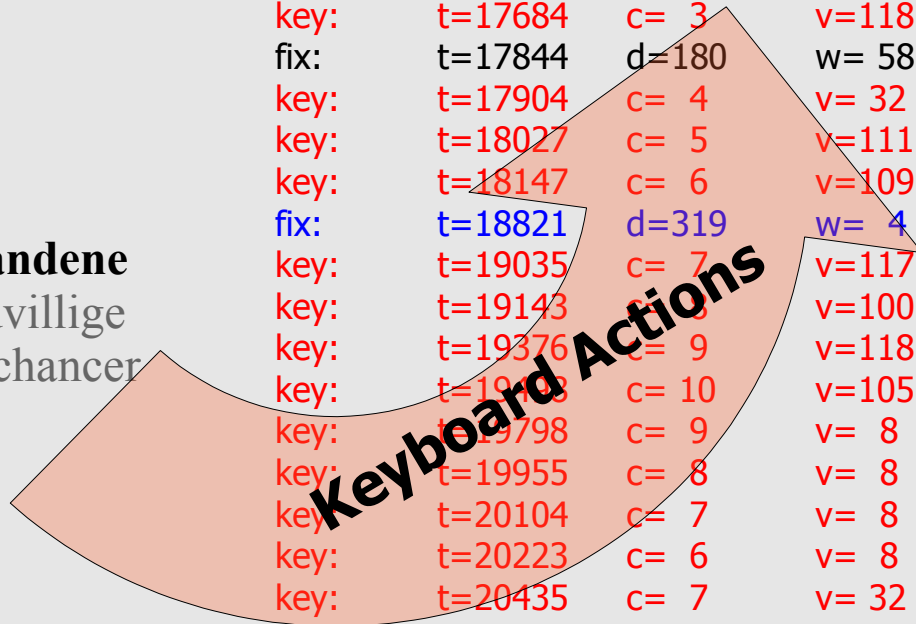
Although **developing countries** are understandably reluctant to compromise ...



Selv om **udviklingslandene** forståeligt nok er modvillige mod at risikere deres chancer ...

Target

fix:	t=12501	d=678	w= 1	developing
fix:	t=13199	d=139	w= 3	are
fix:	t=13358	d=240	w= 4	understandably
fix:	t=13617	d=200	w= 5	reluctant
key:	t=15335	c= 0	v= 83	S
key:	t=15529	c= 1	v=101	e
fix:	t=15771	d=139	w= 0	Although
key:	t=15811	c= 2	v=118	v
key:	t=16149	c= 3	v= 32	
key:	t=16380	c= 2	v= 8	
key:	t=16527	c= 1	v= 8	
key:	t=17362	c= 2	v=108	l
key:	t=17684	c= 3	v=118	v
fix:	t=17844	d=180	w= 58	vulnerable
key:	t=17904	c= 4	v= 32	
key:	t=18027	c= 5	v=111	o
key:	t=18147	c= 6	v=109	m
fix:	t=18821	d=319	w= 4	understandably
key:	t=19035	c= 7	v=117	u
key:	t=19143	c= 8	v=100	d
key:	t=19376	c= 9	v=118	v
key:	t=19498	c= 10	v=105	i
key:	t=19798	c= 9	v= 8	
key:	t=19955	c= 8	v= 8	
key:	t=20104	c= 7	v= 8	
key:	t=20223	c= 6	v= 8	
key:	t=20435	c= 7	v= 32	



History Keyboard Activities on how 'udviklingslandene' emerged from 'developing countries'

User Activity Data

n=48	t=26319	c=25	v=101	e	n=29	t=22885	c=12	v=105	i
n=47	t=26164	c=24	v=110	n	n=28	t=22729	c=11	v=118	v
n=46	t=26058	c=23	v=101	e	n=27	t=22608	c=10	v=8	
n=45	t=25974	c=22	v=100	d	n=26	t=22452	c=11	v=8	
n=44	t=25865	c=21	v=110	n	n=25	t=22313	c=12	v=8	
n=43	t=25740	c=20	v=97 a		n=24	t=21288	c=13	v=107	k
n=42	t=25613	c=19	v=108	l	n=23	t=21068	c=12	v=118	v
n=41	t=25519	c=18	v=115	s	n=22	t=20970	c=11	v=105	i
n=40	t=25329	c=17	v=103	g	n=21	t=20827	c=10	v=100	d
n=39	t=25196	c=16	v=110	n	n=20	t=20649	c=9	v=117	u
n=38	t=25106	c=15	v=105	i	n=19	t=20514	c=8	v=32	
n=37	t=24855	c=14	v=108	l	n=18	t=20414	c=7	v=8	
n=36	t=24731	c=13	v=107	k	n=17	t=20083	c=8	v=8	
n=35	t=24496	c=12	v=8		n=16	t=19934	c=9	v=8	
n=34	t=24130	c=13	v=108	l	n=15	t=19777	c=10	v=8	
n=33	t=23908	c=12	v=8		n=14	t=19477	c=11	v=105	i
n=32	t=23765	c=13	v=8		n=13	t=19355	c=10	v=118	v
n=31	t=23413	c=14	v=107	k	n=12	t=19149	c=9	v=100	d
n=30	t=23118	c=13	v=108	l	n=11	t=19041	c=8	v=117	u

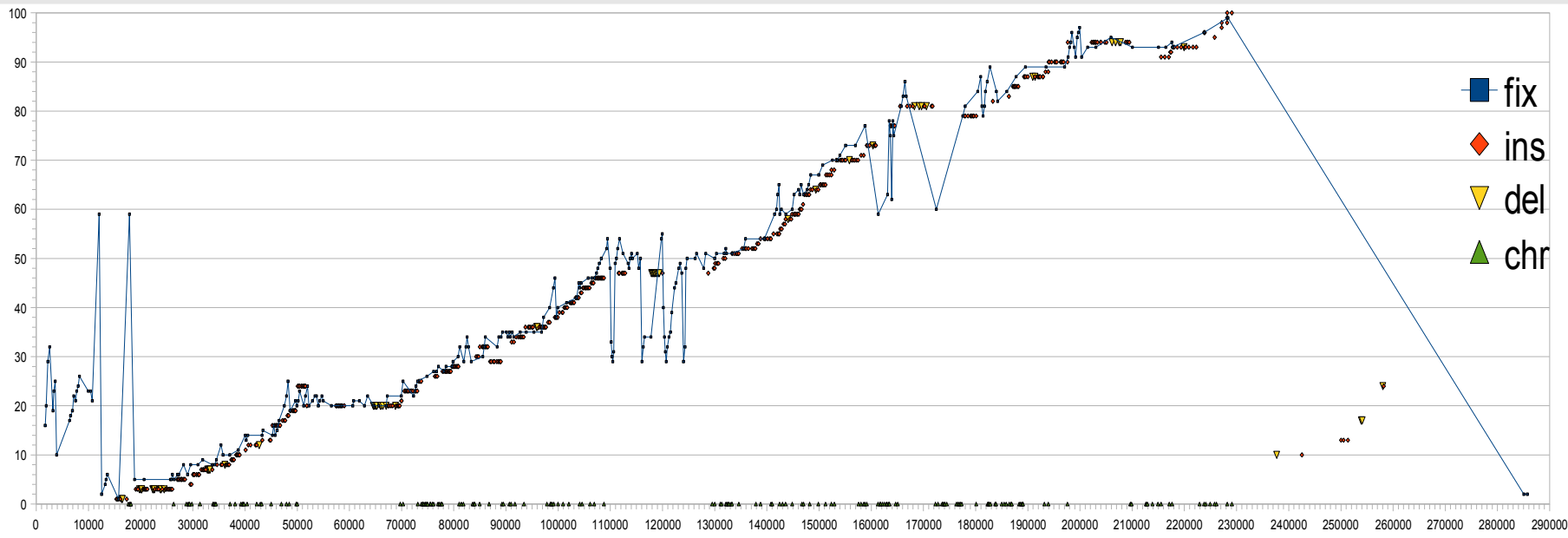
Translation Progression Graph

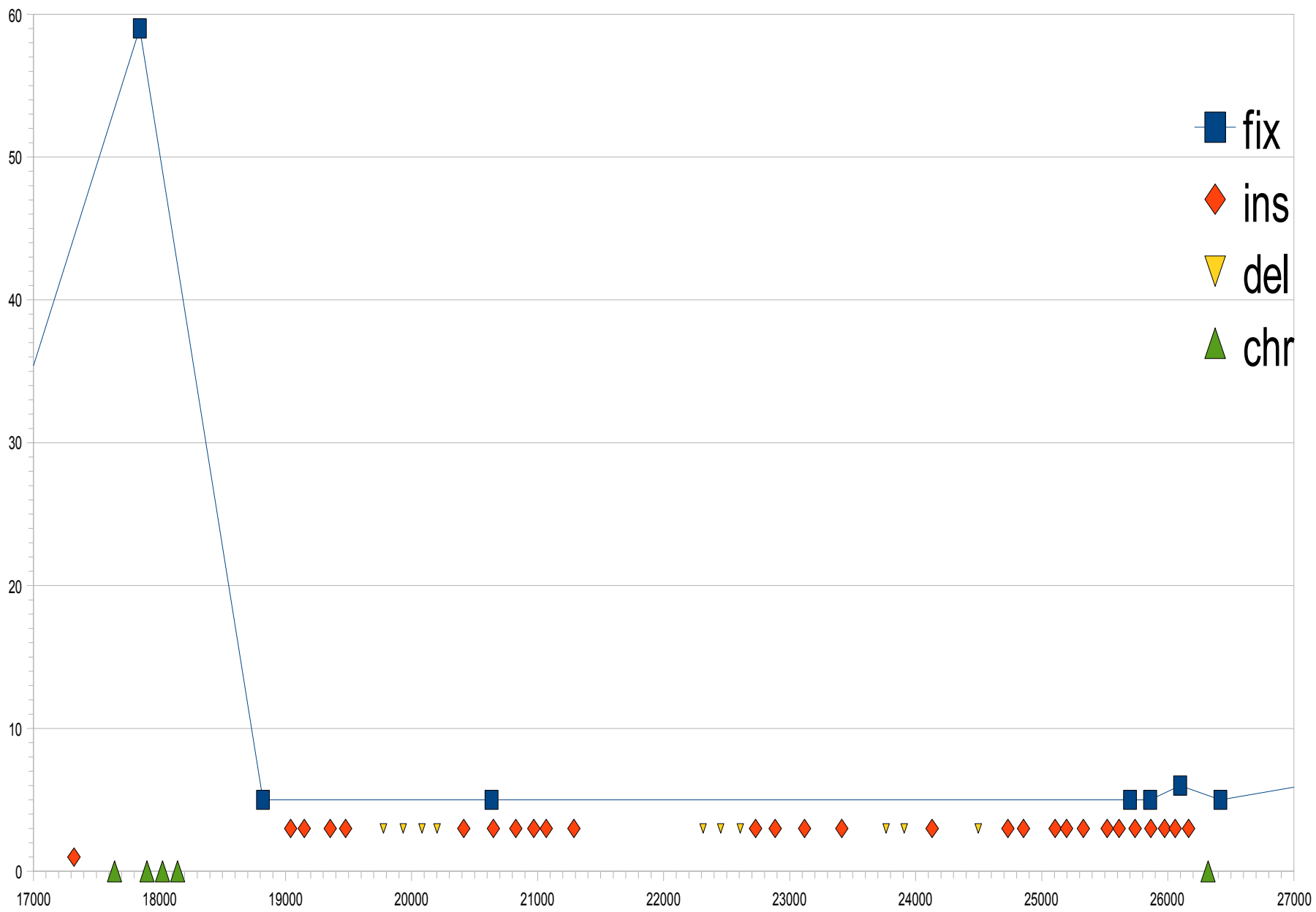
y-axis: 100 words source text

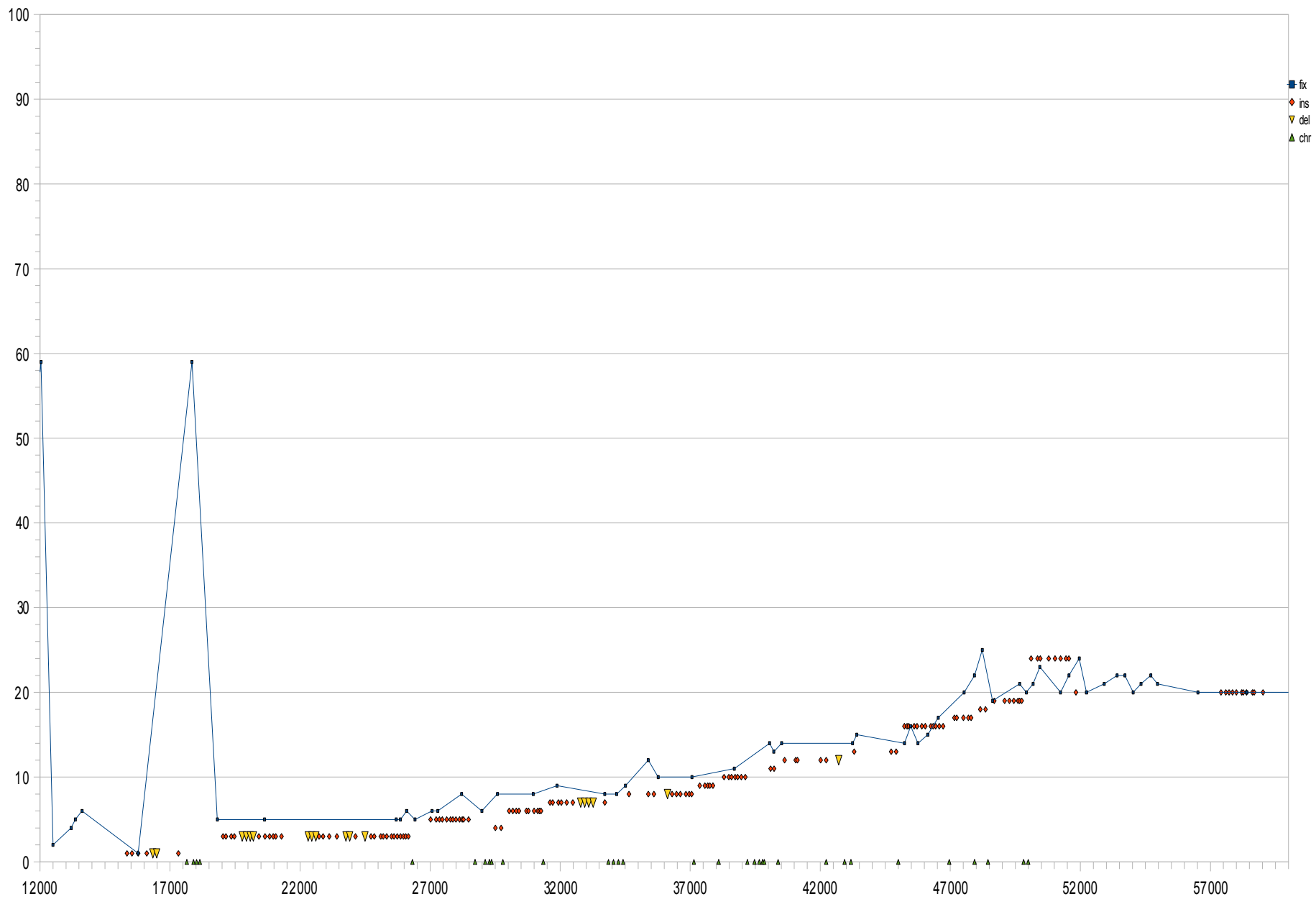
x-axis: 5 mins translation time

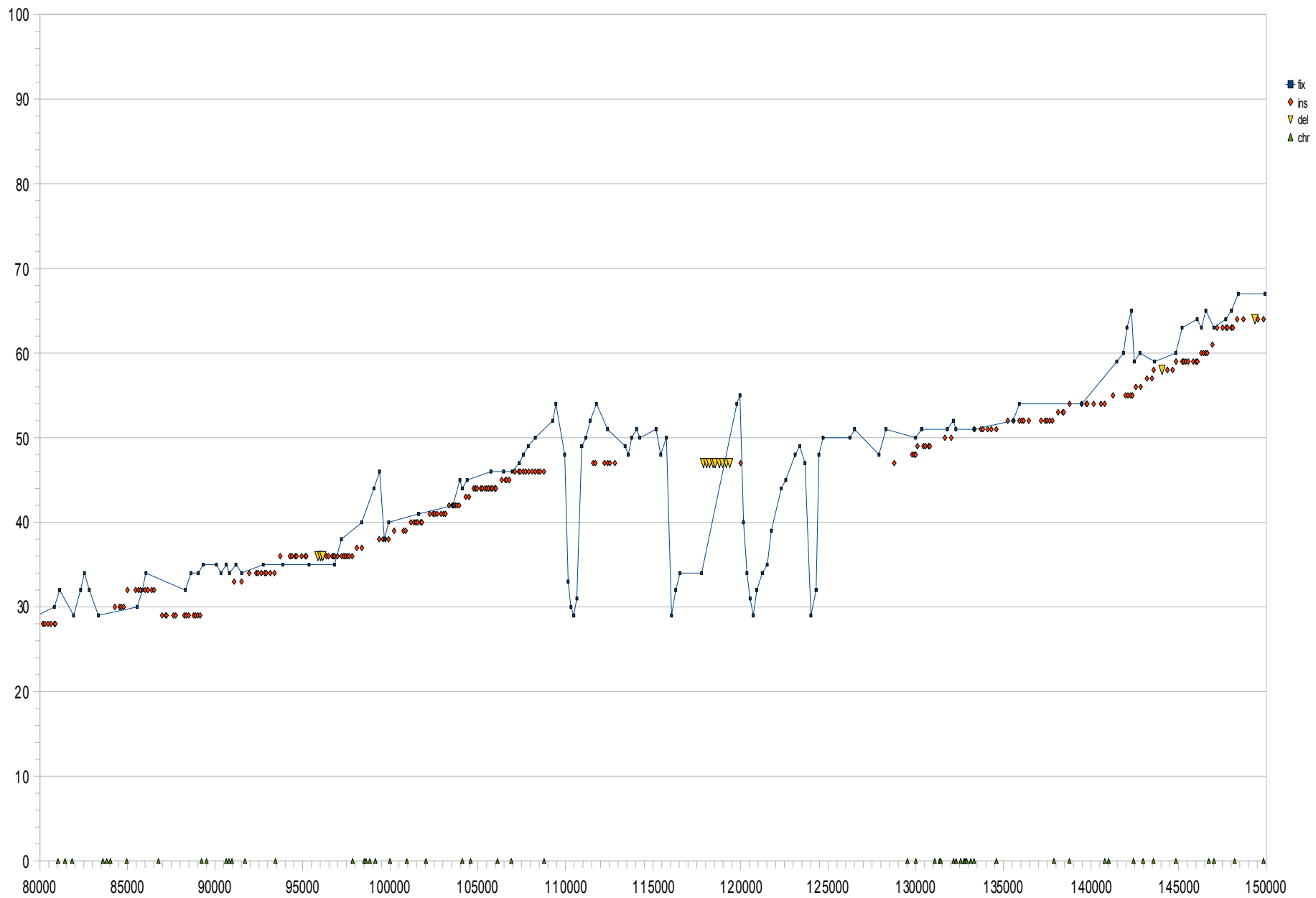
line: fixations on the source text words

dots: keyboard activities: insertions, deletions, unassigned









Compare and correlate UAD

- Different user, same text
 - Compare Fixation patterns
 - Compare pattern of UAD on same linguistic phrases
- Same user, different texts
 - Retrieve text passages with similar fixation patterns
 - Retrieve passages with similar (difficult) typing patterns

Fixation n-grams Reading models

P	#trans	from#	to#	from word	to word
0.25	1	1	3	developing	are
0.75	3	1	2	developing	countries
0.33	1	2	3	countries	are
0.66	2	2	4	countries	understandably
1.00	2	3	4	are	understandably
1.00	8	4	5	understandably	reluctant
0.23	3	5	4	reluctant	understandably
0.15	2	5	5	reluctant	reluctant
0.38	5	5	7	reluctant	compromise
0.07	1	5	1	reluctant	developing
0.15	2	5	6	reluctant	to
0.5	1	6	5	to	reluctant
0.5	1	6	7	to	compromise
0.25	2	7	5	compromise	reluctant
0.25	2	7	7	compromise	compromise
0.37	3	7	8	compromise	their
0.12	1	7	0	compromise	Although
0.5	1	8	7	their	compromise
0.5	1	8	4	their	understandably