Empirical Comparison of PETRARCH-1 and 2 Verb Dictionaries

Philip A. Schrodt

Parus Analytics
Charlottesville, Virginia, USA
schrodt735@gmail.com

Presented at the RIDIR Event Data Project Annual Meeting
University of Texas at Dallas, 11 December 2017

Slides:
`http://eventdata.parusanalytics.com/presentations.html`

History of verb dictionary development

1990 - 2009 : KEDS and TABARI dictionaries developed by countless generations of Kansas political science honors students, some now full professors, working from examples primarily from Reuters and AFP on the Levant, with some additional cases from Balkans, Central Asia and West African

2010 - 2011: WordNet synsets added as part of ICEWS project; organization of verbs into synonym-based classes

2014: TABARI dictionaries used without modification for PETRARCH-1 (P1)

summer 2015: Substantial modifications in the size and form of dictionaries by Caerus Analytics for use in PETRARCH-2 (P2)

Key changes from P1 to P2

- Massive pruning of patterns compared to the P1 dictionaries; remaining patterns are very simple
- Actor tokens for source ($), target (+), compound (%) and actor skipping ($\hat{()}$) are dropped
- Notation for noun phrases (. . .) and prepositional phrases {. . . } is added
- Facility is available for some general transformational rules but only a limited number of these were actually implemented
- Much of the dictionary conversion appears to have been done automatically without human review
- More generally, work on P2 seems to have stopped abruptly with a number of intended features not fully implemented

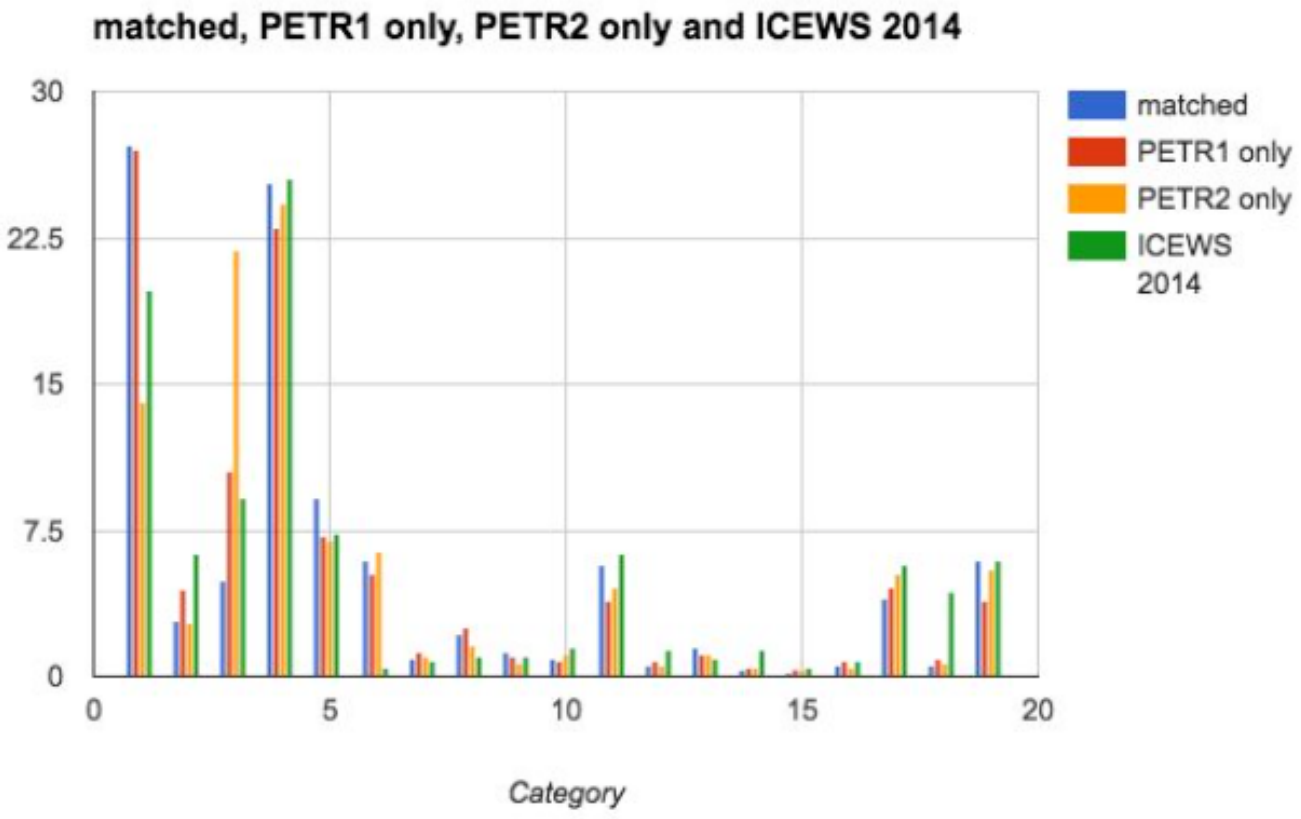
Example of P1 dictionary

```
--- CONFIRM [---] ---
CONFIRM
CONFIDENCE
VALIDATE
SUBSTANTIATE
AFFIRM [051]
REAFFIRM
+SEEM_CONFIDENT [013] # sls 16 Sep 2007
- &CRIME OF $ AIRSPACE BY + &AIRCRAFT * [160]  # CONFIRM <ELH 21 May 2008>
- * COMMITMENT AGAINST THREATS BY + [130]  # REAFFIRM <sls 20 Sep 2007>
- * &MILITARY HOLD POSITION IN + [192]  # CONFIRM <mdc 28 Nov 2006>
- + BEEN * WINNER OF ELECTIONS [060]  # CONFIRM <OY 15 Mar 2006>
- COURT * SENTENCE AGAINST + [173]  # CONFIRM <OY 14 Mar 2006>
- * NOT_INVOLVED_IN_CLASHES [016]  # REAFFIRM <oy 29 Dec 2005>
- ^ * $ CUT &MILITARY TO + [1622]  # CONFIRM <OY 14 Mar 2006>
- $ * ^ NOT EXTRADIT TO + [124]  # CONFIRM <OY 27 Aug 2006>
- * &FIGHT &CRIME CHARGES [1125]  # CONFIRM <lrp 02 Aug 2005>
- * DETERMINATION TO JOIN [031]  # REAFFIRM <OY 28 Aug 2006>
- * ELECTION WILL BE HELD [0341]  # REAFFIRM <OY 17 Mar 2006>
- * WILLING TO TALK TO + [036]  # CONFIRM <ms 11 Apr 2008>
- * SENTENCES &DETENTION [173]  # CONFIRM <jap 18 Jan 2003>
- * &FIGHT &CRIME CHARGE [1125]  # CONFIRM <ab 18 Nov 2005>
- NEITHER * NOR REJECT [011]  # CONFIRM <AI 27 May 2008>
- * BOYCOTT OF MEETING [125]  # CONFIRM <ab 31 Dec 2005>
- * ORDER TO HALT FIRE [010]  # CONFIRM <ab 10 Jul 2003>
- + * &DEFENSE &WEAPON [074]  # CONFIRM <ss 02 Jul 2007>
- * HAD CHOSEN + TANK [062]  # CONFIRM <ss 29 Jun 2007>
- * COMMIT &CEASEFIRE [0356]  # REAFFIRM <oy 13 Dec 2005>
- * PURCHASE &WEAPON [0312]  # AFFIRM <OY 13 Mar 2006>
- * &MILITARY KILLED [010]  # CONFIRM <tb 22 Jan 2007>
- * CLOSE + BORDERS [172]  # CONFIRM <sls 17 Mar 2008>
- AWAIT * OF ARREST [---]  # CONFIRM <OY 25 Jul 2003>
- * COMMITMENT TO + [050]  # AFFIRM <sls 13 Mar 2008>
- * POLICY REFUSING [120]  # REAFFIRM
- * BOYCOTT MEETING [125]  # CONFIRM <ab 31 Dec 2005>
- * CONDITIONS FOR [100]  # REAFFIRM <oy 29 Nov 2005>
- * SECRET PRISONS [015]  # CONFIRM <s;s 16 Nov 2007>
- HANDOVER * BY + [010]  # CONFIRM <ELH 05 May 2008>
- * + WILL_VISIT [036:036]  # CONFIRM <OY 25 Jul 2003>
- * &OPPOSITION [110]  # REAFFIRM
- * WOULD VISIT [036]  # CONFIRM <tony  4/29/91>
- * BOYCOTT OF [163]  # CONFIRM <OY 25 Jul 2003>
- * ALLEGATION [010]  # CONFIRM
- * WOULD_BACK [051]  # CONFIRM <jap 11 Oct 2002>
- * COMMITMENT [050]  # CONFIRM , REAFFIRM <tb 30 May 2007>
```

Example of P2 dictionary

```
---  CONFIRM   [---]  ---
+SEEM_CONFIDENT [013] # sls 16 Sep 2007
AFFIRM [051]
CONFIDENCE
CONFIRM
REAFFIRM
SUBSTANTIATE
VALIDATE
- * ALLEGATION                                [010]        # CONFIRM
- * DISMISSAL                                 [010]        # CONFIRM
- * FIRE                                      [010]        # CONFIRM
- * CONCERN                                   [012]        # REAFFIRM
- * {SECRET PRISONS}                          [015]        # CONFIRM
- * DETERMINATION                             [030]        # REAFFIRM
- * PURCHASE (OF &WEAPON)                     [0312]       # AFFIRM
- * ELECTION                                  [0341]       # REAFFIRM
- * COMMITMENT                                [050]        # AFFIRM
- * COMMITMENT                                [050]        # CONFIRM , REAFFIRM
- * BACKING                                   [051]        # REAFFIRM
- * SUPPORT                                   [051]        # CONFIRM , REAFFIRM
- * WINNER (OF ELECTIONS)                     [060]        # CONFIRM
- * {&DEFENSE &WEAPON}                        [074]        # CONFIRM
- * CONDITIONS                                [100]        # REAFFIRM
- * CLAIM                                     [100]        # REAFFIRM
- * &OPPOSITION                               [110]        # REAFFIRM
- * OBJECTION                                 [110]        # REAFFIRM
- * REJECTION                                 [110]        # REAFFIRM
- * {&FIGHT &CRIME CHARGES}                   [1125]       # CONFIRM
- * BOYCOTT (OF MEETING)                      [125]        # CONFIRM
- * COMMITMENT (AGAINST THREATS)              [130]        # REAFFIRM
- * BOYCOTT                                   [163]        # CONFIRM
- COURT * SENTENCE                            [173]        # CONFIRM
- * ARREST                                    [173]        # CONFIRM
```

Cue category correlations are nonetheless very high



Methodology

- Direct comparison of synonym sets
- Comparison of patterns used by P1 and P2 in coding a corpus consisting of the entire KEDS Levant Reuters (1979-2015) and AFP (2000-2015) stories, and a corpus of about 1.5Gb of news texts from 2014 primarily from the Middle East and Latin America
- Lists of the matching frequencies for all verb classes, verb singletons and patterns were generated and are available on request; I'm largely presenting summary statistics here
- Only matches that generated an event with both a source and target were counted, though these could consist only of agents (i.e. agents did not have to match to an actor)

This was done by adding just a couple of lines of code to P1 and P2 which wrote the matches to a file, then separate Python programs analyzed those files. Programs, as always, are available on request.

Verb synonyms by class

These are essentially identical except for

- Removing some mostly archaic verbs: DECEIVE, SNITCH, TRANSMIT, AIR, COMBUST, INTONE, INTONATE, NAVIGATE, OVERPASS, TRANSVERSE, ACCESS, DISEMBARK, BURST, TRAMPLE, SUFFER, SCOF, IMPORTUNE, POSTULATE, POSIT, INCLUDE.
- Re-assignment to a different class: virtually all of those are from FINISH to CANCEL.
- Differences in the handling words beginning with "RE"

This is very useful since it means that any differences are concentrated in the patterns, not the verbs

Huge differences in the patterns

- Patterns in common: 2593
- Patterns only in PETR1: 7408
- Patterns only in PETR2: 1374

Furthermore, "common" is a high estimate as it is assuming that if the literals—words and synsets—in a pattern are the same, the pattern is equivalent, though that probably is true most of the time.

Frequency distribution for P1

Total matches: 1,090,085

Verbs matched total: 982
matched 2 - 4 times: 81
matched 5 - 9 times: 99
matched >9 times: 750

Patterns matched: 7400
matched 2 - 4 times: 1577
matched 5 - 9 times: 1035
matched >9 times: 3294

Matches on verbs: 516,086
Matches on patterns: 573,999

Frequency distribution for P2

Total matches: 354,153

Verbs matched total: 1167
matched 2 - 4 times: 187
matched 5 - 9 times: 152
matched >9 times: 707

Patterns matched: 3518
matched 2 - 4 times: 953
matched 5 - 9 times: 559
matched >9 times: 1189

Matches on verbs: 242,264
Matches on patterns: 111,889

SIGN class patterns in 90-percentile for P1

Freq	Pattern
2102	% *
559	% * &TREATY
419	* WITH +
419	% * &TREATY WITH +
362	&TREATY *
316	% * &TREATY ON
161	* BETWEEN %
125	* CONTRACT WITH +
115	WILL *
102	WILL % * &TREATY

SIGN class patterns in 90-percentile for P2

Freq	Pattern
988	* &TREATY
143	* CONTRACT
73	* MEMORANDUM
60	* {TRADE &TREATY}
52	* {PEACE &TREATY}
43	* {ECONOMIC &TREATY}

Percentile distributions

Table: PETRARCH-1

Percentile	Verb	Pattern
80.0	224	308
90.0	377	826
95.0	512	1561
99.0	783	3597
99.9	948	6344

Table: PETRARCH-2

Percentile	Verb	Pattern
80.0	228	86
90.0	434	291
95.0	642	622
99.0	1243	1497
99.9	2042	2289

Frequency distribution for P2 auxilary rule usage

Freq	Rule
13322	a (a b Q) SAY = a b Q
1266	a (b . ATTACK) SAY = a b 112
825	a (a b ATTACK) SAY = a b 015
144	a (a b WILL_ATTACK) SAY = a b 138
101	a (b . ATTACK) CLAIM = a b 112
83	a (b . ATTACK) WANT = a b 021
34	a (a b ATTACK) CLAIM = a b 015
26	a (_ b ASSASSINATE) ATTEMPT = a b 185

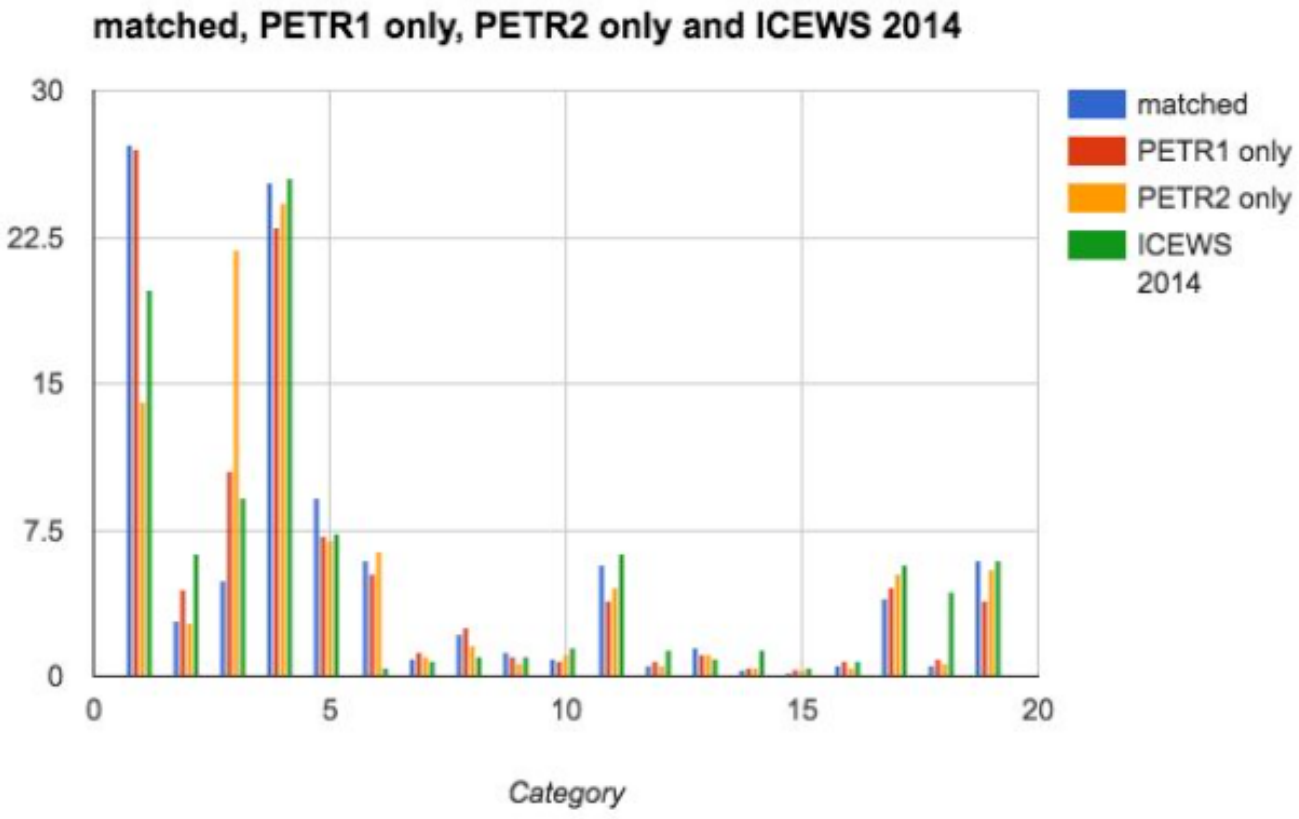
While rules can be *specified* in P2 dictionaries, they are only invoked in about 4.2% of the codings, and just one rule accounts for 3.75% of the matches

Some issues observed in the P2 dictionaries

- WILL and WOULD (along with SCHEDULE) are incongruously in a verb class called - - WILL - -, which means they are coded as 030 ("meeting")
- Some P2 patterns are repeated, occasionally with different codes
- There are null-coded verb classes with verb synonyms but no patterns: these will never code events

I still haven't fully grokked how P2 uses the auxiliary verbs stored in the `meta` structure and some of these may in fact operate as patterns, or they modify codes through the internal `pico` language. More research needs to be done here.

WILL class probably accounts for P2's anomalously high 030 counts



Frequency distribution for P2 auxiliary word usage

Except for the conjunction "and", these may (or may not) be serving the same function as many P1 patterns to modify CAMEO codes

Freq	Word
24857	and
18076	HAS
7490	IS
7363	HAVE
4953	MEET
4021	WAS
3538	HAD
2039	WERE
1129	TAKE
715	COULD
653	SAY
610	GO
607	MAY
518	URGED
507	GIVE
494	SHOULD

Additional high-frequency P2 auxiliary words

Freq	Word
424	LEAVE
373	TO
350	MUST
327	SEEK
259	DID
239	TRIED
213	PAY
207	BUY
189	WITHDRAW
166	FIGHT
163	BREAK
159	TELL
152	EXPEL
120	STRIKE
118	MIGHT
115	LEAD
114	SUBMIT
105	DEAL
103	PLAN

Verb classes where P1 codes proportionately more than P2

```
 5.40  AID  (0.95%, 0.41%)
 5.70  KILL  (3.76%, 3.19%)
 5.97  BOOST  (1.01%, 0.41%)
 6.23  ESTIMATE  (0.73%, 0.11%)
 6.27  HOPE  (0.88%, 0.26%)
 6.50  ALLEGE  (1.11%, 0.46%)
 6.51  EMPHASIZE  (0.95%, 0.30%)
 6.83  DIE  (1.12%, 0.44%)
 6.95  RE_ADMIT  (0.70%, 0.00%)
 7.31  DEMAND  (1.98%, 1.25%)
 7.38  TARGET  (0.95%, 0.21%)
 9.16  RELEASE  (1.49%, 0.57%)
 9.30  SUGGEST  (1.31%, 0.38%)
10.06  INTEND  (1.51%, 0.51%)
10.16  HURT  (2.51%, 1.49%)
10.40  CALLED_FOR  (1.04%, 0.00%)
11.05  CONSIDER  (1.44%, 0.33%)
13.38  ADD  (1.34%, 0.00%)
15.22  DISCLOSE  (1.83%, 0.31%)
17.02  PREDICT  (2.18%, 0.47%)
17.93  GRANT  (3.49%, 1.70%)
18.21  ALLOW  (2.69%, 0.87%)
19.29  OBSERVE  (2.33%, 0.40%)
30.88  APPREHEND  (5.82%, 2.73%)
32.15  OBTAIN  (3.37%, 0.15%)
258.63  EXPLAIN_VERBAL  (40.23%, 14.37%)
```

Verb classes where P2 codes proportionately more than P1

```
-81.26  WILL  (0.00%, 8.13%)
-16.25  CALL  (1.95%, 3.57%)
-14.89  START  (0.45%, 1.94%)
-11.41  BLAME  (2.72%, 3.86%)
-10.51  ACCLAIM  (0.66%, 1.71%)
-8.52  WARN  (0.22%, 1.07%)
-6.85  REFUSE  (0.27%, 0.95%)
-6.07  SELECT  (0.00%, 0.61%)
-5.18  ASK  (0.31%, 0.83%)
-4.92  DENOUNCE  (0.56%, 1.05%)
-4.68  VISIT  (0.81%, 1.27%)
-4.49  CONFER  (1.88%, 2.33%)
-4.25  MEET_PEOPLE  (2.35%, 2.78%)
-3.93  DESTROY  (0.40%, 0.80%)
-3.90  PHYSICAL_CONFLICT  (1.62%, 2.01%)
-3.84  SHOOT  (0.41%, 0.80%)
-3.84  APPEAL  (0.10%, 0.49%)
-3.63  CANCEL  (0.00%, 0.36%)
-3.34  PREPARE  (0.11%, 0.44%)
-3.33  SEND  (0.12%, 0.45%)
-3.09  EXPLOSIVE_DESTRUCTION  (1.26%, 1.5
-2.94  SEARCH  (0.11%, 0.41%)
-2.91  ARRIVE  (1.62%, 1.91%)
-2.87  ASSEMBLE  (0.65%, 0.93%)
-2.29  RESTART  (0.07%, 0.30%)
-2.25  MARCH  (0.23%, 0.46%)
-2.15  REACH  (0.00%, 0.21%)
-2.10  VOW  (0.00%, 0.21%)
-2.02  SOUND  (0.00%, 0.20%)
```

Small sample of classes where P1 and P2 have similar frequencies

```
0.04  PATROL  (0.01%, 0.01%)
0.04  DEFUSE  (0.00%, 0.00%)
0.04  DENIGRATE  (0.01%, 0.00%)
0.04  TERRORIZE  (0.00%, 0.00%)
0.04  CURSE  (0.01%, 0.00%)
0.04  OVERTHREW  (0.01%, 0.00%)
0.04  PARDON  (0.01%, 0.01%)
0.04  SPY  (0.01%, 0.00%)
0.05  SAVAGE  (0.01%, 0.00%)
0.05  INTRUDE  (0.01%, 0.00%)
0.05  RECAPTURE  (0.02%, 0.02%)
0.06  EXTRADITE  (0.04%, 0.04%)
0.06  IMMIGRATE  (0.01%, 0.00%)
0.07  SUPPRESS  (0.06%, 0.05%)
0.07  SUE  (0.05%, 0.04%)
0.08  FINISH  (0.09%, 0.08%)
0.08  LEGALISE  (0.01%, 0.00%)
0.08  BLACKLIST  (0.01%, 0.00%)
0.09  VIOLATE  (0.03%, 0.02%)
0.09  ASSASSINATE  (0.05%, 0.05%
0.10  PARADE  (0.02%, 0.01%)
0.11  FORGIVE  (0.01%, 0.00%)
0.11  ALLAY  (0.02%, 0.01%)
0.11  VANDALISE  (0.02%, 0.01%)
0.12  ENDANGER  (0.02%, 0.01%)
0.12  BESIEGE  (0.04%, 0.03%)
0.12  INVITE  (0.23%, 0.21%)
0.13  POUND  (0.14%, 0.12%)
0.14  DISGRACE  (0.02%, 0.01%)
0.14  CONSPIRE  (0.01%, 0.00%)
0.14  REFUTE  (0.01%, 0.00%)
0.15  NOTIFY  (0.02%, 0.01%)
0.15  FLATTEN  (0.02%, 0.01%)
```

What might be done: major development issues

- Presumably we are focusing on UD-PETRARCH (UD-P) now, not P2
- CAMEO or PLOVER coding?
  - If PLOVER, do we also try to implement the mode and context codes where this is easy?
  - PLOVER also drops the single most-frequent category, comments
- Get frequencies on additional corpora, which is easy
- This methodology can easily be applied to the Spanish and Arabic dictionaries

What might be done: improving the dictionaries

- Do the NP and PP patterns in UD-P/P2 dictionaries actually pick up the P1 token-based patterns or is a separate facility needed for these?
- Review the UD-P dictionaries to delete any weirdness such as duplicate patterns
- Are the high-frequency patterns coding correctly based on a human (crowd-sourced used spAcy's PRODIGY system?) review of results?
- Should the P2 rule set be implemented in UD-P (if it hasn't been already) and then expanded to encompass a large number of verbs?

Final thought: Isn't this totally doomed anyway?

Should we not simply welcome our new Tensorflow-based deep learning overlords? See "IARPA." Well, maybe, maybe not...

- I've reviewed several neural-network-based event coding papers recently and none are remotely close to production-level accuracy
- The BBN ACCENT/SERIF coder used for ICEWS and its derivatives remains dictionary-based
- Any deep-learning approach will require a very large number of gold-standard training examples
  - We still don't have these: there is no event-data equivalent to the massive parallel corpora used to enhance deep-learning translation algorithms
  - GSRs could also be used to augment and validate dictionaries

Final thought: Isn't this totally doomed anyway? (pt 2)

- Natural language is different from images, and event data coding—which must be very precise on grammatically-structured sentences which typically have little redundant information—is very different than the general problem of language translation, where "close enough" works
  - Consequently patterns + syntax may contain a very high density of the required classification information
  - The P1 dictionaries incorporate about twenty years of human coding
  - There is little cost to including low-frequency patterns provided they are accurate
- There is still no widely accepted "killer app" for event data but the approach does not seem to be going away
- We may want a hybrid approach for PLOVER with sentence-level dictionaries used to assign events, article-level classifiers (SVM might be sufficient) used to assign mode and context

Thank you

Email:
schrodt735@gmail.com

Slides:
`http://eventdata.parusanalytics.com/presentations.html`