NewsReader

Building structured event indexes
of large volumes of financial and economic data
for making decisions
FP7-2012-ICT-315404

Event Detection

Itziar Aldabe / German Rigau / Aitor Soroa
IXA group, UPV/EHU
http://ixa.si.ehu.es

Project Objectives

- Event Detection in English, Dutch, Spanish and Italian
  - events in terms of *who* did *what when* and *where*
  - relations between events
  - factual / non-factual or speculative
  - provenance: who tells what and when
- Narrative schemas and storylines over longer periods of time
- Event reasoning: richness, coherence, relevance
- Large-scale processing, storage and retrieval of events (integrate new with old)

Project Objectives

- **Event Detection in English, Dutch, Spanish and Italian**
  - **events in terms of *who* did *what* *when* and *where***
  - **relations between events**
  - **factual / non-factual or speculative**
  - **provenance: who tells what and when**
- Narrative schemas and storylines over longer periods of time
- Event reasoning: richness, coherence, relevance
- Large-scale processing, storage and retrieval of events (integrate new with old)

Event Detection: Main goals

- **Events**, **participants**, their **roles** in text
- **Time** and **place** expressions
- **Attribution** of the events (aka *Authority* and *Factuality*)
- Standard **output** and **scaling**

- English (UPV/EHU, VUA, FBK)
- Spanish (UPV/EHU)
- Dutch (VUA)
- Italian (FBK)

Event Detection: Example

Mariano Rajoy announced yesterday in Madrid that the budget cuts will continue next year.

Event Detection: Example

Mariano Rajoy **announced** yesterday in Madrid that the budget cuts will continue next year.

event1: announced

Event Detection: Example

Mariano Rajoy **announced** yesterday in Madrid that the budget cuts will continue next year.

event1: announced

who: Mariano Rajoy

what: that the budget cuts will continue next year

when: yesterday

where: in Madrid

Event Detection: Example

Mariano Rajoy announced yesterday in Madrid that the budget cuts will **continue** next year.

event1: announced

- who: Mariano Rajoy
- what: that the budget cuts will continue next year
- when: yesterday
- where: in Madrid

event3: continue

- what: the budget cuts
- when: next year

Event Detection: Example

Mariano Rajoy announced yesterday in Madrid that the budget **cuts** will continue next year.

event1: announced
- who: Mariano Rajoy
- what: that the budget cuts will continue next year
- when: yesterday
- where: in Madrid

event2: cuts
- what: budget

event3: continue
- what: the budget cuts
- when: next year

Event Detection: Example

Mariano Rajoy **announced** yesterday in Madrid that the budget cuts will continue next year.

event1: announced

who: Mariano Rajoy

what: that the budget cuts will continue next year

when: yesterday

where: in Madrid

Event Detection: Example

Mariano Rajoy **announced** yesterday in Madrid that the budget cuts will continue next year.

event1: announced

who: Mariano Rajoy NERC (PER)

what: that the budget cuts will continue next year

when: yesterday

where: in Madrid NERC (LOC)

Event Detection: Example

Mariano Rajoy **announced** yesterday in Madrid that the budget cuts will continue next year.

event1: announced

who: Mariano Rajoy

- http://dbpedia.org/resource/Mariano_Rajoy NED

what: that the budget cuts will continue next year

when: yesterday

where: in Madrid

- http://dbpedia.org/resource/Madrid NED

Event Detection: Example

Bush announced yesterday in Madrid that the budget cuts will continue next year.

- http://dbpedia.org/page/George_H._W._Bush
- http://dbpedia.org/page/George_W._Bush
- http://dbpedia.org/page/Jeb_Bush

Event Detection: Example

Mariano Rajoy **announced** yesterday in Madrid that the budget cuts will continue next year.

event1: announced

- who: Mariano Rajoy
- what: that the budget cuts will continue next year
- when: yesterday
  - 2015-03-04 TIMEX3
- where: in Madrid

Event Detection: Example

Mariano Rajoy **announced** yesterday in Madrid that the budget cuts will continue next year.

event1: announced => announce.01

who: Mariano Rajoy

- Arg0-PAG: announcer (vnrole: 37.7-1-Agent, 48.1.2-Agent)

what: that the budget cuts will continue next year

- Arg1-PPT: utterance (vnrole: 37.7-1-Topic, 48.1.2-Theme)

when: yesterday

- AM-TMP

where: in Madrid

- AM-LOC

SRL + Predicate Matrix

Event Detection: NLP tools

Mariano Rajoy announced yesterday in Madrid that the budget cuts will continue next year.

NERC NED COREF EVENTS PARTICIPANTS TIME ...

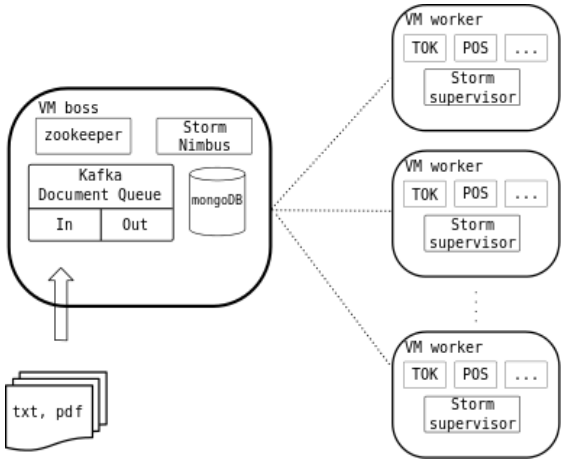
FACTUALITY ATTRIBUTION OPINION ...

Outline

- System architecture & NAF
- EN, SP, IT and NL pipelines
- Evaluation of EN pipeline
- Cross-lingual interoperability

Architecture

- A system composed of many virtual machines (VMs) distributed across one or more physical machines
- Two types of VMs:
  - *boss VM*: The main/manager node
  - *worker VM*: The VMs that perform the processing. They contain all the NLP modules



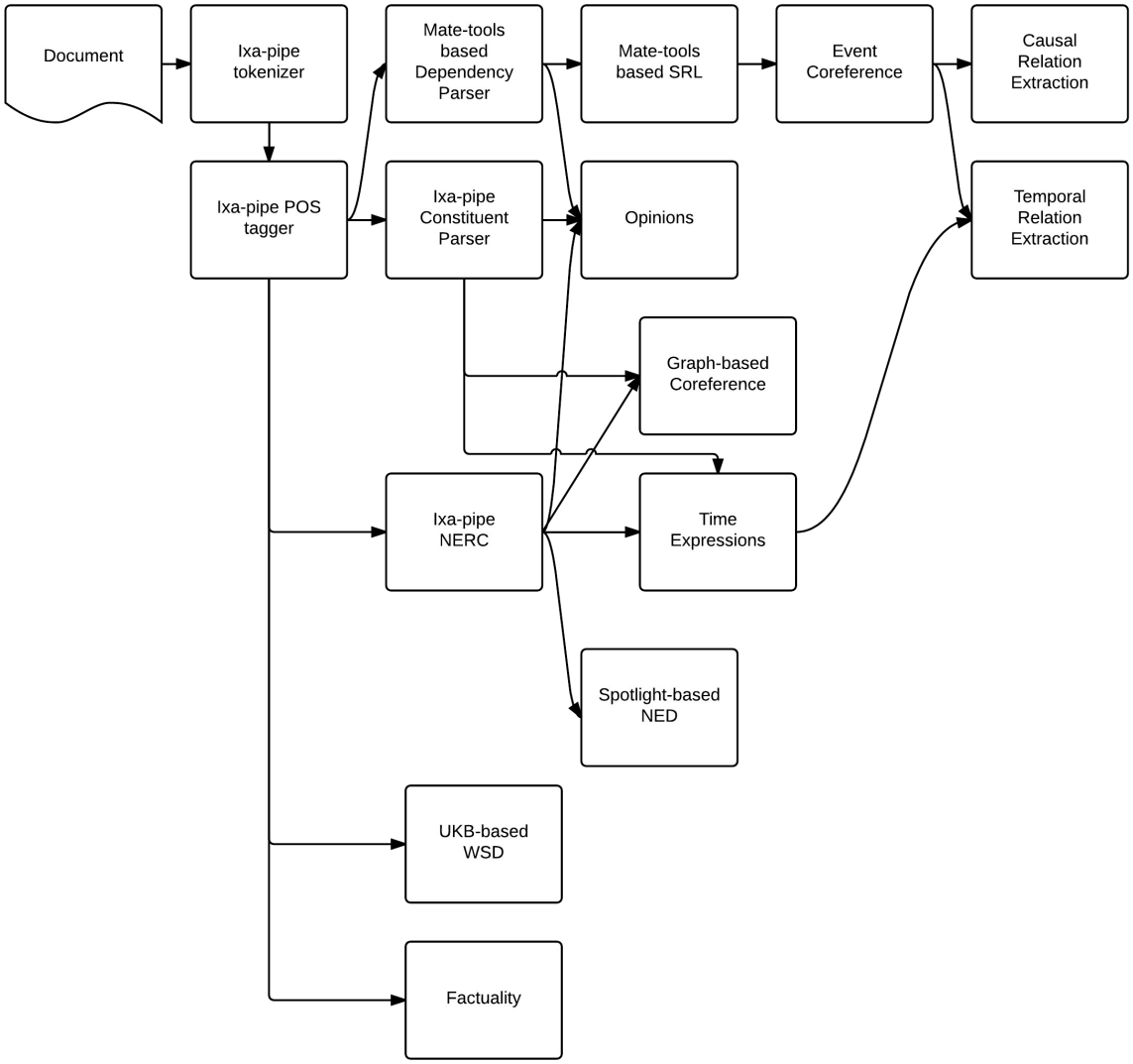
Standards for inter-operability

- NAF
  - Basic format for inter-document NLP analysis
  - Stand-off XML, multi-layered annotation format
  - Allows parallel processing
  - Covers many linguistic levels
  - All NLP modules read and write NAF

English generic pipeline v2.1

1) NAF libraries (EHU, VUA)
2, 3, 4) Ixa-pipes (EHU): tok, pos, nerc
5, 6) ukb (EHU), vua-svm-wsd (VUA)
7) ixa-pipe-srl (mate tools + predicate-matrix) :
   dependency parsing + SRL (N+V) + event-classification
8) ixa-pipe-parse (EHU) : constituent parsing
9) corefgraph (EHU): nominal coreference
10) ixa-pipe-ned (DBpedia spotlight) : Entity Linking
11) fbk-timepro (FBK) : timex3
12) vua-factuality (VUA)
13) vua-opinion-miner (VUA)
14) vua-event-coref (VUA) : intra-document event coreference
15) fbk-temprel (FBK)
16) fbk-causalrel (FBK)

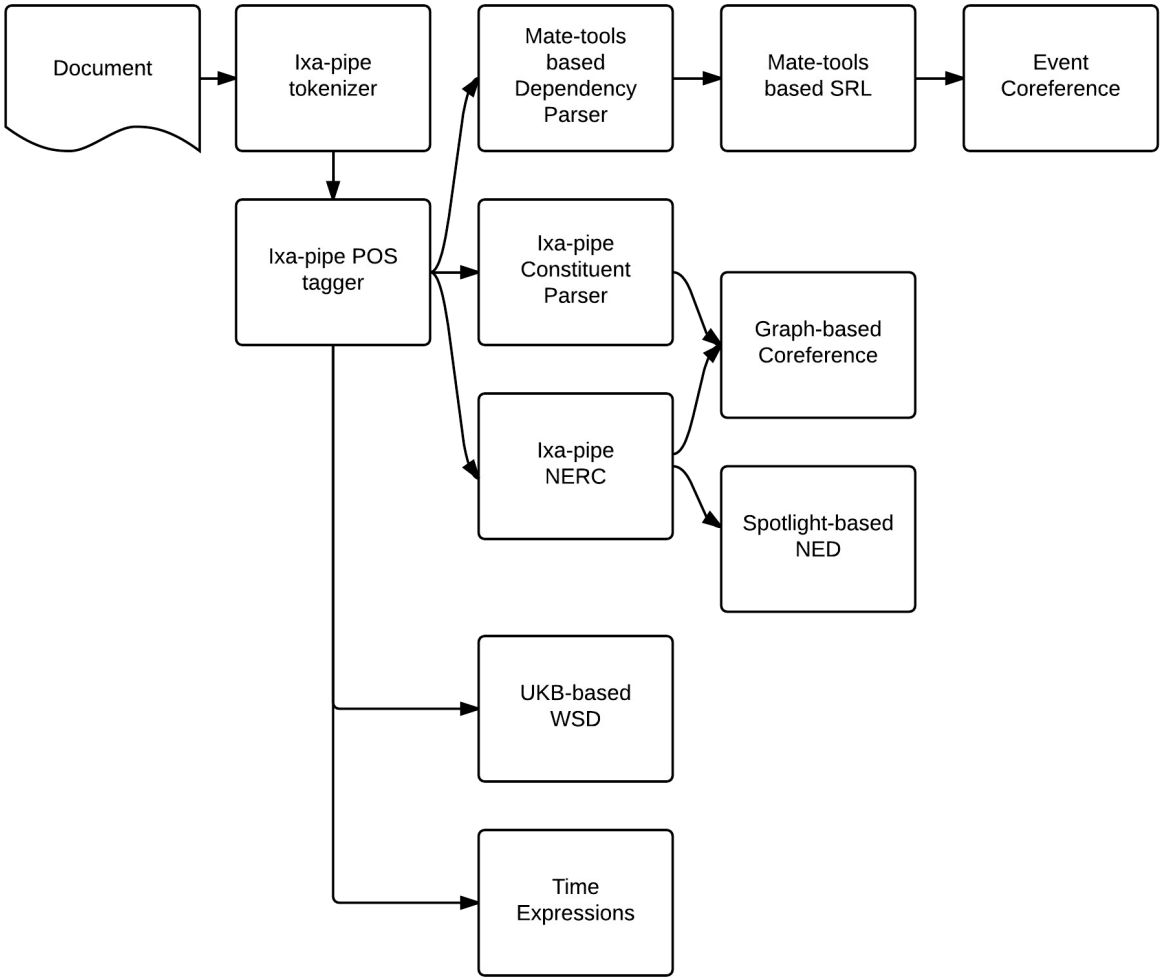
English generic pipeline v2.1



Spanish generic pipeline v2.0

1) NAF libraries (EHU, VUA)
2, 3, 4) ixa-pipes (EHU): tok, pos, nerc
5) ixa-pipe-wsd (ukb)
6) ixa-pipe-srl (mate tools + multilingual predicate matrix) :
dependency parsing + SRL (N+V) + event-classification
7) ixa-pipe-parse (EHU) : constituent parsing
8) corefgraph (EHU): nominal coreference
9) ixa-pipe-ned (DBpedia spotlight) : cross-lingual Entity Linking
10) vua-event-coref (VUA) : intra-document event coreference
11) ixa-heidel-time (EHU) : timex3

Spanish generic pipeline v2.0



Dutch generic pipeline v2.0

1) NAF libraries (EHU, VUA)
2,3) ixa-pipes (EHU): tok, nerc
4) vua-alpino (VUA): pos, parsing
5) vua-svm-wsd (VUA)
6) vua-srl (VUA)
7) vua-ontotagging (VUA) : cross-lingual Predicate-matrix tagging
8) vua-framenet-classifier (VUA)
9) corefgraph (EHU): nominal coreference
10) ixa-pipe-ned (DBpedia spotlight) : cross-lingual Entity Linking
12) vua-opinion-miner (VUA)
13) vua-event-coref (VUA) : intra-document event coreference
14) vua-heidel-time (VUA)

Italian generic pipeline v2.0

1) fbk-tokenpro (FBK)
2) fbk-morphopro (FBK)
3) fbk-tagpro (FBK)
4) fbk-lemapro (FBK)
5) fbk-entitypro (FBK)
6) fbk-chunkpro (FBK)
7) fbk-timepro (FBK)
8) fbk-syntaxpro (FBK)
9) fbk-eventpro (FBK)
10) fbk-temprel (FBK)
11) fbk-factpro (FBK)

Large-scale processing

- LN car company news (EN) : ~ 6M
  - 1st year: 63K
  - 2nd year: 1.3M articles
- TechCrunch (EN) : 43K articles
- Kiem (EN) : 212K documents
- Dutch House of Representatives (NL) : 1M documents
- Europarl (EN, ES) : ~19K EN, ~19K ES
- WikiNews (EN, ES, IT, NL) :
  - 19K EN, 8K IT, 7K ES and 1K NL
- ECB+ (EN) : 984 articles

Evaluation of pipelines

- Intra-document Benchmarking
  - On standard datasets (EN)
  - WikiNews (EN, ES, IT, NL) :
    - 19K English, 8K Italian, 7K Spanish and 1K Dutch
    - 120 documents (EN => ES, IT, NL)
    - Timelines (SemEval-2015)
  - NERC, NED, Nominal Coref (EN)
  - SRL, Event Detection, Event Coref (EN)
  - Temporal Processing (EN, IT)
- *Cross-document Benchmarking (EN, ES, IT, NL)*
  - *...*

Evaluation of pipelines

- NERC evaluation (ixa-pipes-nerc)

Table 6: NERC CoNLL 2003 testb results.

	Precision	Recall	F1
Newsreader (ixa-pipe-nerc)	91.64	90.21	**90.92**
Stanford NER	-	-	88.08
Ratinov et al. (2009)	-	-	90.57
Passos et al. (2014)	-	-	90.90

Table 7: NERC Intra-document Benchmarking with Wikinews.

System	mention extent	Precision	Recall	F1
Newsreader (ixa-pipe-nerc combined)	Inner phrase-based	62.46	76.26	**68.67**
Stanford NER (all english distsim)	Inner phrase-based	63.53	68.21	65.79
Illinois NER (CoNLL 2003)	Inner phrase-based	57.19	69.44	62.72
Newsreader (ixa-pipe-nerc combined)	Inner token-based	72.37	79.43	**75.74**
Stanford NER (all english distsim)	Inner token-based	77.14	71.77	74.36
Illinois NER (CoNLL 2003)	Inner token-based	69.24	72.93	71.04
Newsreader (ixa-pipe-nerc combined)	Outer phrase-based	53.30	68.24	**59.85**
Stanford NER (all english distsim)	Outer phrase-based	52.86	59.51	55.99
Illinois NER (CoNLL 2003)	Outer phrase-based	47.53	60.52	53.24
Newsreader (ixa-pipe-nerc combined)	Outer token-based	73.50	67.20	**70.21**
Stanford NER (all english distsim)	Outer token-based	78.22	60.63	68.31
Illinois NER (CoNLL 2003)	Outer token-based	70.26	61.66	65.68

Evaluation of pipelines

- NED evaluation (DBpedia-spotlight)
  - Standard datasets
    - AIDA 79.67 precision 75.94 recall
    - TAC KBP 79.77 precision 60.68 recall
  - Wikinews

Corpus	Precision	Recall	Gold	System-NERC	System-NED
Airbus	54.79%	53.87%	292	316	292
GM	60.98%	36.83%	*505*	321	305
Stock	55.20%	30.51%	331	194	183
Total	57.31%	39.45%	1133	968	780

Table 8: Performance of the NED module on the WikiNews dataset

Evaluation of pipelines

- Nominal Coreference (Corefgraph)

Table 20: Multi-sieve Pass and CoNLL 2011 dev-auto Evaluation

System	MUC	B^3	CEAF	BLANC	CONLL 2011 F1
Stanford	59.6	68.3	45.5	73.0	59.3
Newsreader (Corefgraph)	51.0	67.2	43.4	69.7	54.8

Table 22: Wikinews Nominal Coreference Evaluation

System	MUC	B^3	CEAF	CONLL 2011 F1
Newsreader (Corefgraph)	19.70	18.34	18.96	19.00

Evaluation of pipelines

- Semantic Role Labelling (mate)

Labeled precision	(19137 + 10036) / (22467 + 10818)	87.65%
Labeled recall	(19137 + 10036) / (24748 + 10818)	82.02%
Labeled F1		84.74%
Unlabeled precision	(20697 + 10818) / (22467 + 10818)	94.68%
Unlabeled recall	(20697 + 10818) / (24748 + 10818)	88.61%
Unlabeled F1		91.55%

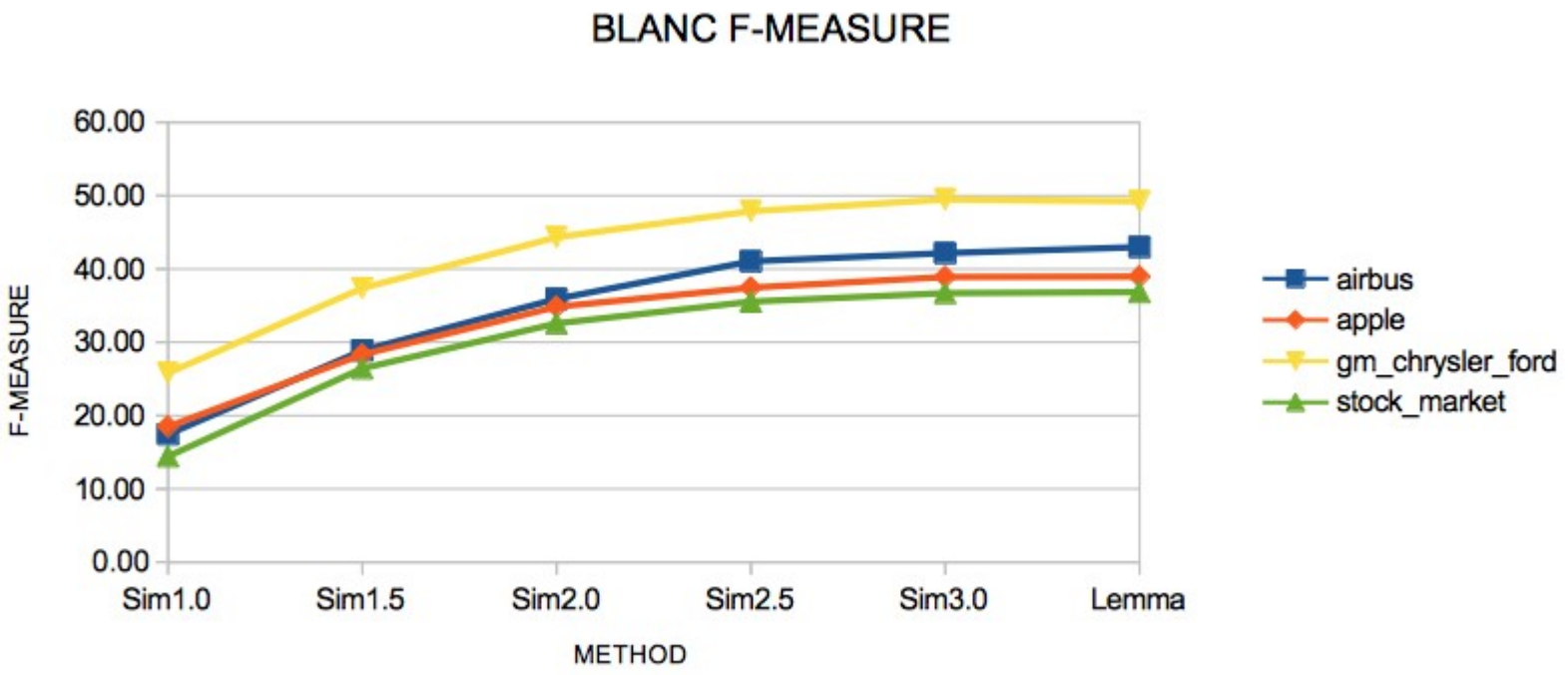
Table 23: Performance of MATE on the English dataset of CoNLL-2009

Labeled precision	(3288 + 1186) / (4154 + 1186)	83.78%
Labeled recall	(3288 + 1186) / (5314 + 1338)	67.26%
Labeled F1		74.62%
Unlabeled precision	(4154 + 1186) / (4154 + 1186)	100.00%
Unlabeled recall	(4154 + 1186) / (5314 + 1338)	80.28%
Unlabeled F1		89.06%

Table 26: Performance of MATE over the GS annotations of WikiNews with the full span from NAF.

Evaluation of pipelines

- Event Coreference



Evaluation of pipelines

- Time detection (TimePro)

	recognition			*classification*	*normalization*
	recall	precision	F1-score	F1-score type	F1-score value
strict match					
Apple	0.805	0.968	0.879	0.831	0.715
Airbus, Boeing	0.761	0.909	0.828	0.817	0.793
GM, Chrysler, Ford	0.714	0.905	0.798	0.782	0.613
Stock market	0.636	0.881	0.738	0.692	0.58
Micro-average	0.72	0.914	**0.805**	0.773	**0.664**
relaxed match					
Apple	0.841	1	0.913	0.856	0.721
Airbus, Boeing	0.848	0.974	0.907	0.86	0.79
GM, Chrysler, Ford	0.789	1	0.882	0.849	0.647
Stock market	0.709	0.982	0.823	0.762	0.646
Micro-average	0.787	0.99	**0.877**	0.827	**0.692**

Table 27: TimePro performance

Evaluation of pipelines

- Time Relations (TimeRelPro)

	recall	precision	F1-score
strict match			
Apple	0.174	0.361	0.235
Airbus, Boeing	0.145	0.319	0.199
GM, Chrysler, Ford	0.194	0.474	0.275
Stock market	0.081	0.141	0.103
Micro-average	0.154	0.325	0.209
relaxed match			
Apple	0.177	0.366	0.238
Airbus, Boeing	0.155	0.341	0.213
GM, Chrysler, Ford	0.196	0.479	0.278
Stock market	0.084	0.146	0.106
Micro-average	0.158	0.333	0.214
temporal awareness			
Apple	0.201	0.387	**0.265**
Airbus, Boeing	0.157	0.328	**0.212**
GM, Chrysler, Ford	0.221	0.484	**0.303**
Stock market	0.094	0.155	**0.117**
Micro-average	0.173	0.339	**0.229**

Table 28: TempRelPro performance

SemEval 2015 Task 4

- TimeLine: Cross-document Event Ordering

- iTunes

1	2003	11778-3-launch	11778-4-launch
2	2007	11778-4-pass	
3	2008-01	11778-7-hold	
4	2008-02	11778-2-pass	11778-5-pass
4	2008-02	11778-3-accounts_for	

Evaluation of pipelines

- SemEval 2015 Task 4
- TimeLine: Cross-Document Event Ordering

Participant	CORPUS1 F1	CORPUS2 F1	CORPUS3 F1	TOTAL F1 score	TOTAL Precision	TOTAL Recall
Track A						
SPINOZAVU_1	4.07	5.31	0.42	3.15	7.95	1.96
SPINOZAVU_2	2.67	0.62	0.00	1.05	8.16	0.56
WHUNLP_1	9.42	5.97	7.26	**7.85**	14.59	5.37
NWR-TE	*15.99*	*12.48*	*12.42*	*13.88*	*15.94*	*12.29*
SubTrack A						
SPINOZAVU_1	1.20	1.70	2.08	**1.69**	6.70	0.97
SPINOZAVU_2	0.00	0.92	0.00	0.27	13.04	0.14
WHUNLP_1	*3.91*	*0.87*	*2.21*	*2.59*	*9.29*	*1.51*
NWR-TE	*5.49*	*7.59*	*7.51*	*6.94*	*8.41*	*5.90*
Track B						
GPLSIUA_1	22.35	19.28	33.59	**25.36**	21.73	30.46
GPLSIUA_2	20.47	16.17	29.90	22.66	20.08	26.00
HeidelToul_1	19.62	7.25	20.37	17.03	20.11	14.76
HeidelToul_2	16.50	10.94	25.89	18.34	13.58	28.23
SubTrack B						
GPLSIUA_1	18.35	20.48	32.08	**23.15**	18.90	29.85
GPLSIUA_2	15.93	14.44	27.48	19.18	16.19	23.52
HeidelToul_1	12.23	14.78	16.11	14.42	19.58	11.42
HeidelToul_2	13.24	15.88	21.99	16.67	12.18	26.41

Table 30: Official results of TimeLine task. The results in italics were not submitted officially to SemEval-2015 Task 4.

Evaluation summary

- Most modules perform as SOA
- NERC beyond SOA (standard & out of domain)
- Nominal coref slightly below SOA
- Timelines beyond SOA (by far)

- Most complete generic pipelines (EN, ES, NL, IT)
- Cross-lingual semantic interoperable

- Very robust processing

Cross-lingual interoperability

- **Named entities**
  - **Cross-lingual links** from DBpedia
- **Events**
  - **Predicate Matrix**
  - **Interoperable** event models
    - VN, FN, PB, WN (SUMO, etc.), ESO
  - Extending to **nominal** predicates (EN)
  - **Cross-lingual** PM (EN, ES, NL, IT)
- **Time**
  - **Cross-lingual** time normalization
- **Concepts**
  - **Cross-lingual** wordnets
  - MCR, MultiWordNet, OMW

Cross-lingual interoperability: Named entities

Mariano Rajoy announced yesterday in Madrid that the budget cuts will continue next year.

```
    <entity id="e1" type="PERSON">
      <references>
        <!--Mariano Rajoy-->
        <span>
          <target id="t1" />
          <target id="t2" />
        </span>
      </references>
      <externalReferences>
        <externalRef resource="spotlight_v1" reference="http://dbpedia.org/resource/Mariano_Rajoy"
confidence="1.0" reftype="en" />
      </externalReferences>
    </entity>
```

Cross-lingual interoperability: Named entities

Mariano Rajoy anunció ayer en Madrid que los recortes continuarán el próximo año.

```
    <entity id="e1" type="PERSON">
      <references>
        <!--Mariano Rajoy-->
        <span>
          <target id="t1" />
          <target id="t2" />
        </span>
      </references>
      <externalReferences>
        <externalRef resource="spotlight_v1" reference="http://es.dbpedia.org/resource/Mariano_Rajoy"
confidence="0.9999999" reftype="es">
          <externalRef resource="wikipedia-db-esEn"
reference="http://dbpedia.org/resource/Mariano_Rajoy" confidence="0.9999999" reftype="en" />
        </externalRef>
```

Cross-lingual interoperability: Events

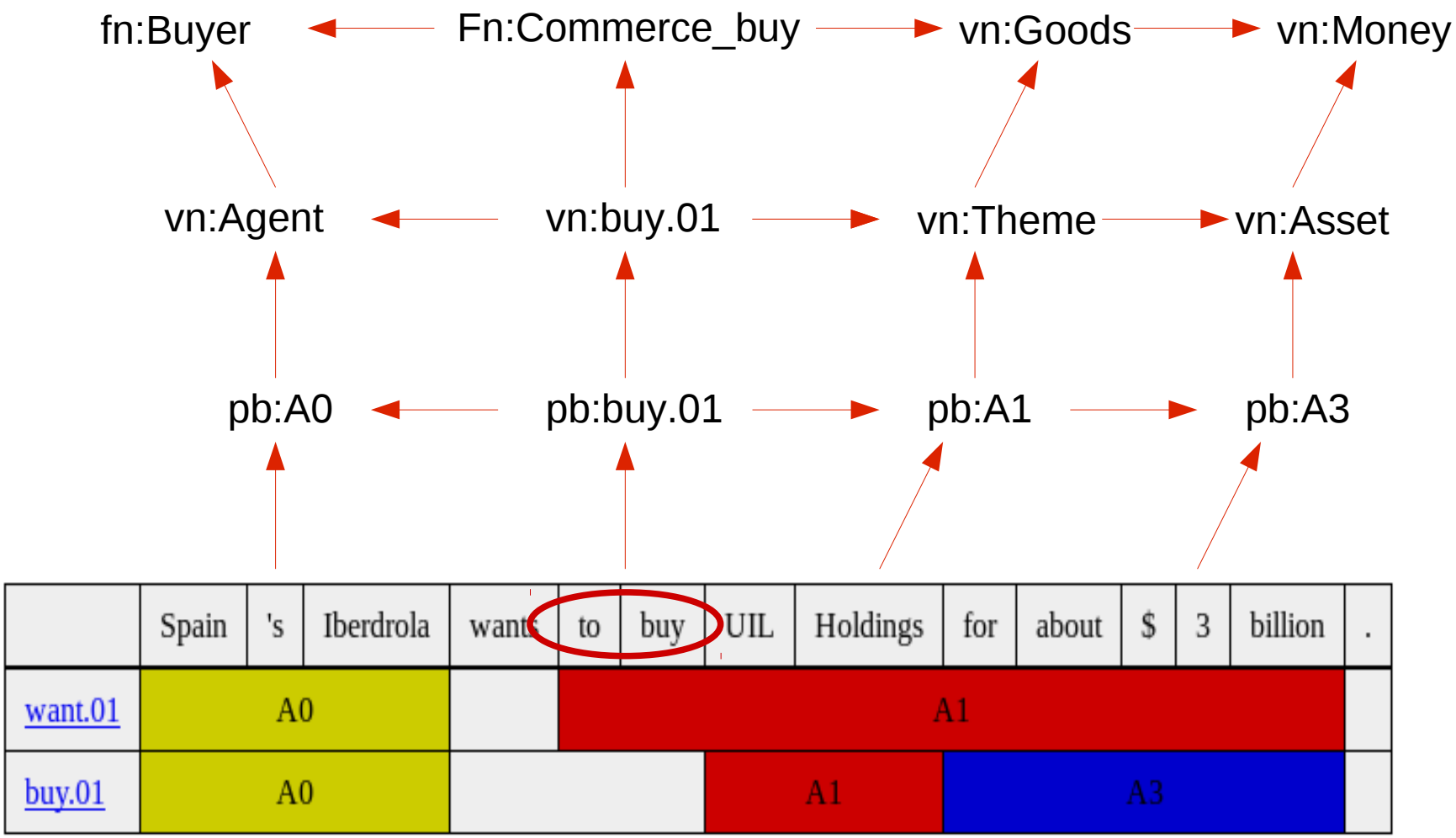
- Predicate Matrix is a new lexical resource resulting from the integration of multiple sources of predicate information including FrameNet, VerbNet, PropBank and WordNet.

- http://adimen.si.ehu.es/web/PredicateMatrix

Predicate Matrix

	Spain	's	Iberdrola	wants	to	buy	UIL	Holdings	for	about	$	3	billion	.
want.01	A0				A1									
buy.01	A0						A1			A3				

Predicate Matrix



Cross-lingual interoperability: Events

Mariano Rajoy **announced** yesterday in Madrid that the budget cuts will continue next year.

```
<srl>
  <!--t3 announced : A0[t1 Mariano] AM-TMP[t4 yesterday] AM-LOC[t5 in] A1[t7 that]-->
  <predicate id="pr1">
    <!--announced-->
    <span>
      <target id="t3" />
    </span>
    <externalReferences>
      <externalRef resource="PropBank" reference="announce.01" />
      <externalRef resource="VerbNet" reference="reflexive_appearance-48.1.2" />
      <externalRef resource="VerbNet" reference="say-37.7" />
      <externalRef resource="VerbNet" reference="say-37.7-1" />
      <externalRef resource="FrameNet" reference="Statement" />
      <externalRef resource="PropBank" reference="announce.01" />
      <externalRef resource="EventType" reference="communication" />
      <externalRef resource="WordNet" reference="ili-30-00974367-v" />
      <externalRef resource="WordNet" reference="ili-30-00975427-v" />
    </externalReferences>
    <role id="rl1" semRole="A0">
      <!--Mariano Rajoy-->
      <span>
        <target id="t1" />
        <target id="t2" head="yes" />
      </span>
      <externalReferences>
        <externalRef resource="VerbNet" reference="reflexive_appearance-48.1.2@Agent" />
        <externalRef resource="VerbNet" reference="say-37.7@Agent" />
        <externalRef resource="FrameNet" reference="Statement@Speaker" />
        <externalRef resource="PropBank" reference="announce.01@0" />
      </externalReferences>
    </role>
```

Cross-lingual interoperability: Events

Mariano Rajoy **anunció** ayer en Madrid que los recortes continuarán el próximo año.

```
<srl>
  <!--t3 anunci? : arg0[t1 Mariano] argM[t4 ayer] argM[t5 en] arg1[t7 que]-->
  <predicate id="pr1">
    <!--anunci?-->
    <span>
      <target id="t3" />
    </span>
    <externalReferences>
      <externalRef resource="AnCora" reference="anunciar.1.benefactive" />
      <externalRef resource="VerbNet" reference="reflexive_appearance-48.1.2" />
      <externalRef resource="VerbNet" reference="say-37.7" />
      <externalRef resource="VerbNet" reference="say-37.7-1" />
      <externalRef resource="FrameNet" reference="Statement" />
      <externalRef resource="PropBank" reference="announce.01" />
      <externalRef resource="EventType" reference="communication" />
      <externalRef resource="WordNet" reference="ili-30-00974367-v" />
      <externalRef resource="WordNet" reference="ili-30-00975427-v" />
    </externalReferences>
    <role id="rl1" semRole="arg0">
      <!--Mariano Rajoy-->
      <span>
        <target id="t1" head="yes" />
        <target id="t2" />
      </span>
      <externalReferences>
        <externalRef resource="VerbNet" reference="reflexive_appearance-48.1.2@Agent" />
        <externalRef resource="VerbNet" reference="say-37.7@Agent" />
        <externalRef resource="FrameNet" reference="Statement@Speaker" />
        <externalRef resource="PropBank" reference="announce.01@0" />
      </externalReferences>
    </role>
```

Cross-lingual interoperability: Time

Mariano Rajoy announced **yesterday** in Madrid that the budget cuts will continue **next year**.

```
<timeExpressions>
  <timex3 id="tmx0" type="DATE" value="2015-03-05" />
  <timex3 id="tmx1" type="DATE" value="2015-03-04">
    <!--yesterday-->
    <span>
      <target id="w4" />
    </span>
  </timex3>
  <timex3 id="tmx2" type="DATE" value="2016">
    <!--next year-->
    <span>
      <target id="w13" />
      <target id="w14" />
    </span>
  </timex3>
</timeExpressions>
```

Cross-lingual interoperability: Time

Mariano Rajoy anunció **ayer** en Madrid que los recortes continuarán **el próximo año**.

```
<timeExpressions>
  <timex3 id="tx1" type="DATE" value="2015-03-05" />
  <timex3 id="tx2" type="DATE" value="2015-03-04">
    <!--ayer-->
    <span>
      <target id="w4" />
    </span>
  </timex3>
  <timex3 id="tx3" type="DATE" value="2016">
    <!--el pr?ximo a?o-->
    <span>
      <target id="w11" />
      <target id="w12" />
      <target id="w13" />
    </span>
  </timex3>
</timeExpressions>
```

Cross-lingual interoperability: Concepts

A militant **Korean** nationalist has slashed the face of the US ambassador to South Korea at a breakfast meeting in Seoul.

```
<!--Korean-->
<term id="t3" type="open" lemma="korean" pos="G" morphofeat="JJ">
  <span>
    <target id="w3" />
  </span>
  <externalReferences>
    <externalRef resource="wn30g.bin64" reference="ili-30-02967791-a" confidence="1.0" />
    <externalRef resource="WordNet-3.0" reference="ili-30-02967791-a" confidence="1.0" />
  </externalReferences>
</term>
```

Cross-lingual interoperability: Concepts

Un militante nacionalista **coreano** ha rajado la cara del embajador estadounidense en Corea del Sur en un desayuno de trabajo en Seúl.

```
<!--coreano-->
<term id="t4" type="open" lemma="coreano" pos="G" morphofeat="AQ0MS0">
  <span>
    <target id="w4" />
  </span>
  <externalReferences>
    <externalRef resource="wn30sp.bin64" reference="ili-30-02967791-a" confidence="1.0" />
  </externalReferences>
</term>
```

Open source modules and VMs

- Open source code (>30 modules)
  - http://github.com/newsreader
- Virtual Machines for clusters (automatic deployment)
  http://github.com/ixa-ehu/vmc-from-scratch

Event Detection – Main results

- EN, ES, NL and IT **generic** pipelines
- EN pipeline adapted to different **domains**
- **Cross-lingual interoperable** semantic interpretation:
  - DBpedia
  - Cross-lingual PM
  - Time normalization
  - Cross-lingual senses
- **Open source** modules and VMs
- EN, ES, NL and IT **new large-scale processing**
- **Evaluation** of NWR pipelines
  - On standard datasets (EN)
  - On wikinews & Timelines - SemEval 2015 (EN)
  - On Eventi task - Evalita 2014 (IT)

Event Detection – Future work

- EN, ES, NL, IT pipelines adapted to **financial domain**

- Improved EN, ES, NL and IT **generic** pipelines
- **Cross-lingual interoperable** semantic interpretation
- **Open source** modules and VMs
- EN, ES, NL and IT **new large-scale processing**
- **Evaluation** of NWR pipelines
  - Monolingual & Cross-lingual

- Improving **interacting** tasks
- Automatic **domain adaptation** techniques

Event Detection – Future work

- Improved EN generic pipelines
  - Topic classification – JEX
  - Wikification – dbpedia-spotlight
  - NED-reranker
  - Frame-reranker
  - **Implicit arguments => Egoitz**

NewsReader

Building structured event indexes
of large volumes of financial and economic data
for making decisions
FP7-2012-ICT-315404

WP4 Event Detection

Itziar Aldabe / German Rigau / Aitor Soroa
IXA group, UPV/EHU
http://ixa.si.ehu.es