

Going Hybrid with Time

Francisco Costa and António Branco



University of Lisbon
{fcosta, antonio.branco}@di.fc.ul.pt

DELPH-IN Meeting 2012
July 3, 2012

- ① Introduction
- ② Tense
- ③ Aspect
- ④ Hybrid Approach
- ⑤ Conclusions

- 1 Introduction**
- 2 Tense
- 3 Aspect
- 4 Hybrid Approach
- 5 Conclusions

Goals

- Extracting temporal information from text in a linguistically motivated way
- Implementation of tense in the Portuguese grammar LXGram

Motivation

- Natural language describes time in a complex way
- Possibly too complex to get it completely right initially
- But implementing it enables us to find the real difficulties

Challenges

- Ambiguity of tense
- Tense comes with aspect
- Aspect comes with aspect shift/aspect coercion
- Temporal information depends on extra-linguistic factors
 - Calendar systems
 - Logical properties of temporal relations
 - Pragmatics
- Lexical aspect: more information in the lexicon

Approach

- Ignore lexical aspect. . . in the lexicon
 - Hard to annotate reliably
 - Would require annotation of a large part of the lexicon
- But account for the compositional side of aspect
- Hybrid approach: use an external component to handle extra-linguistic information

The implementation builds on much of the literature on tense and aspect

- [Reichenbach 1947], [Davidson 1967], [Vendler 1967], [Partee 1973], [Dowty 1979], [Comrie 1985], [Pustejovsky 1991]
- [de Swart 1998, de Swart 2000], [Bonami 2002], [Goss-Grubbs 2005], [Flouraki 2006]

In this presentation

- A small overview of some of the implementation of tense and aspect in LXGram, skipping over several issues
- Combination of LXGram with an external temporal component

- 1 Introduction
- 2 Tense**
- 3 Aspect
- 4 Hybrid Approach
- 5 Conclusions

Ambiguity at two levels

- Surface \leftrightarrow grammatical tense
E.g.: English *put* can be simple present or simple past
Portuguese *corremos* “we run/ran”: past or present
- Grammatical tense \leftrightarrow semantics
E.g.: simple present can have future readings:
the train leaves tomorrow

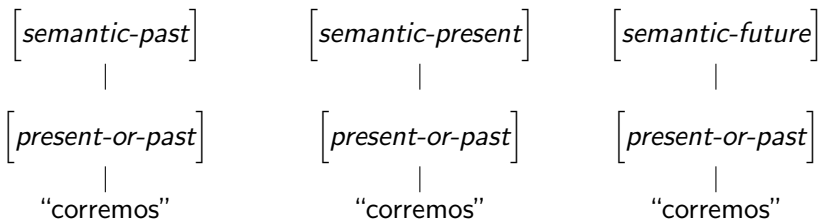
Ideally we underspecify ambiguity

Two possible solutions:

- Cross-classification of semantics and morphology
- **Two sets of lexical rules**

Solution with two levels of lexical rules for tense

- Lowest set of rules maps between form and morphological features
- Second level maps between morphological features and semantics



The rules for **semantic tense** constrain the **grammatical/morphological tense** of their input in different ways:

- $\left[\begin{array}{l} \textit{semantic-past} \\ \text{DTR|MORPH|VERBAL-M|M-TENSE } \textit{past} \end{array} \right]$
- $\left[\begin{array}{l} \textit{semantic-present} \\ \text{DTR|MORPH|VERBAL-M|M-TENSE } \textit{present} \end{array} \right]$
- $\left[\begin{array}{l} \textit{semantic-future} \\ \text{DTR|MORPH|VERBAL-M|M-TENSE } \textit{present-or-future} \end{array} \right]$

The temporal semantic representations are added in this second level

Semantic representation of tense

- Event variables, like the other DELPH-IN HPSGs
- An *at* relation between event variables and a temporal index, representing the event time
- Temporal indices *t* stand for time intervals
- Speech time/utterance time: subtype *now*
- Temporal relations between temporal indices

Example

O gato é gordo "The cat is fat"

$_o_q(x_1, _gato_n(x_1), _gordo_a(e, x_1) \wedge at(e, t) \wedge include(t, now))$

Semantic content of the various tenses: past tenses I

- Imperfective and perfective grammatical aspect
- E.g. *chovia* (imperfective) / *choveu* (perfective) “it rained”
- Imperfective past (*pretérito imperfeito*)
 - The situation held at some point in the past
 - It may still hold in the present
 - $at(e, t_1) \wedge overlap(t_1, t_2) \wedge before(t_2, now)$
 - Temporal modifiers introduce an *overlap* temporal relation with the event time: $\dots \wedge overlap(t_3, t_1)$
 - *Chovia* “It rained (it used to rain)”:
 $at(e, t_1) \wedge overlap(t_1, t_2) \wedge before(t_2, now) \wedge _chover_v(e)$

Semantic content of the various tenses: past tenses II

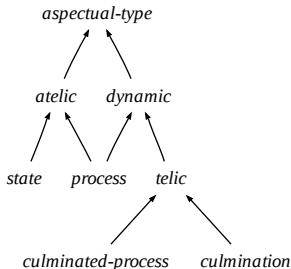
- Perfective past (*pretérito perfeito*)
 - The situation held at some point in the past
 - It no longer holds in the present
 - $at(e, t_1) \wedge before(t_1, now)$
 - Temporal modifiers introduce an *include-or-equal* temporal relation with the event time: $\dots \wedge include-or-equal(t_2, t_1)$
 - *Ontem choveu* “Yesterday it rained”:
 $at(e, t_1) \wedge before(t_1, now) \wedge _chover_v(e)$

Semantic content of the various tenses: present and future

- Present
 - Always imperfective
 - $at(e, t_1) \wedge include(t_1, now)$
- Future and future of past/conditional
 - In Portuguese, no overt perfectivity distinctions
 - The Portuguese grammatical future tense is ambiguous between imperfective and perfective readings
 - Telic readings in sentences with stative predicates
 - Habitual readings in sentences with dynamic predicates
 - We should have two semantic tenses for each
 - In practice that doubles the number of parses
 - The same problem occurs with future readings of grammatical present
 - Also future of past readings of *pretérito imperfeito*
 - Compromise: only $at(e, t_1) \wedge after(t_1, now)$ readings, but with no aspectual constraints (which are described next)

- 1 Introduction
- 2 Tense
- 3 Aspect**
- 4 Hybrid Approach
- 5 Conclusions

Aspectual types similar to [Dowty 1979] and [Vendler 1967]



Tense constrains aspect at the clausal level

- Imperfective tenses constrain the clause to be a state
- Perfective tenses constrain it to be telic
- Other elements can also constrain aspect at various levels (e.g. durational adverbials, VP selecting verbs, etc.)

Aspectual type is often modelled by typing event variables with types that encode aspectual class

In LXGram we model it with Boolean features under events

- `ASPECTUAL-TYPE|CULMINATION` is $+$ for telic situations (culminations and culminated processes)
- `ASPECTUAL-TYPE|PROCESS` is $+$ for processes and culminated processes
- `ASPECTUAL-TYPE|STATE` is $+$ for states

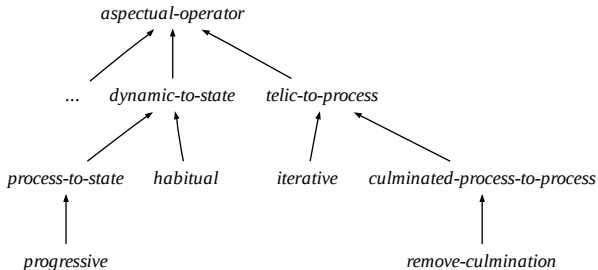
These features are used instead of the aspectual types so that we do not lose parses should there be a bug in the grammar (due to unification failures)

Lexical aspect is left underspecified, but it could be made explicit

Syntactic constraints on aspect are implemented

Aspect shift is represented with aspectual operators

- Aspectual operators are functions between situations of different types
- They are organized in a type hierarchy, like [Bonami 2002]



Aspect sensitive elements introduce these operators in the semantic representations

Problem

- [Bonami 2002] uses implicit aspectual operators (IAO) to model coercion
 - $imparfait-infl \rightarrow \left[\begin{array}{l} \text{KEY } \boxed{1} \\ \text{RELS } \langle \boxed{1}, \text{imp-rel}, (\text{IAO-rel}) \rangle \end{array} \right]$
 - These are introduced in the semantic representations iff there is a clash
 - E.g. the imperfective past constrains its clause to be a state; the IAO-rel is introduced in the MRS representation iff it is not
 - The IAOs model the shifts in meaning that occur in such cases
- But we can't underspecify the size of `RELS` in the LKB
- Adopted solution: always include an aspectual operator, but assume that it can stand for the identity function
 - If an aspectual operator relates two situations of the same aspectual type, it is assumed to be this identity function

Example with present tense: *Chove* “It rains”

- Present: $at(e_1, t_1) \wedge include(t_1, now) \wedge$
 $aspectual-operator(e_1\{state+\}, e_2, _chover_v(e_2))$
- Present tense constrains the whole clause to be a state, because it is an imperfective tense
 - Hence the constraint on the feature STATE of e_1
- Lexical aspect not encoded; if it were:
 - *Chover* “rain” is lexically a process
 - The feature PROCESS of e_2 would be constrained to be positive
 - The aspectual operator would be a function from processes to states, e.g. the habitual operator, which is what we want

Example with present tense II: *O gato é gordo* “The cat is fat”

- Present: $\dots at(e_1, t_1) \wedge include(t_1, now) \wedge aspectual-operator(e_1\{state+\}, e_2, -gordo_a(e_2, x_1)) \dots$
- Ser “to be” is lexically a state:
 - The feature STATE of e_2 would be constrained to be positive
 - Both e_1 and e_2 are states
 - Therefore the aspectual operator is the identity function
 - No aspect coercion
 - No shift in meaning due to aspect coercion

Tenses revisited (examples with *chover* “rain”):

- Imperfective past: *Chovia* “It rained”
 - $at(e_1, t_1) \wedge overlap(t_1, t_2) \wedge before(t_2, now) \wedge aspectual-operator(e_1\{state+\}, e_2, _chover_v(e_2))$
 - Possible interpretation: “it used to rain” (habitual reading)
- Perfective past: *Choveu* “It rained”
 - $at(e_1, t_1) \wedge before(t_1, now) \wedge aspectual-operator(e_1\{culmination+\}, e_2, _chover_v(e_2))$
 - *Chover* “to rain” is lexically a process
 - The operator is a function from processes to culminations or culminated processes
 - One such possibility consists in adding a culmination to the original process, making it a culminated process, which is the intended reading (“it rained for a while and then it stopped”).

Interaction between tense and aspect: the progressive

- Imperfective past
 - *Estava a chover/Estava chovendo*
“It was raining”
 - $at(e_1, t_1) \wedge overlap(t_1, t_2) \wedge before(t_2, now) \wedge$
 $aspectual-operator(e_1\{state+\}, e_2\{state+\},$
 $progressive(e_2, e_3\{process+\},$
 $aspectual-operator(e_3, e_4, _chover_v(e_4))))$
 - Intended meaning:
 - $e_1 = e_2$ (both are states) and $e_3 = e_4$ (both are processes)
- Perfective past
 - *Esteve a chover/Esteve chovendo*
“It was raining (and then it stopped)”
 - $at(e_1, t_1) \wedge before(t_1, now) \wedge$
 $aspectual-operator(e_1\{culmination+\}, e_2\{state+\},$
 $progressive(e_2, e_3\{process+\},$
 $aspectual-operator(e_3, e_4, _chover_v(e_4))))$

- 1 Introduction
- 2 Tense
- 3 Aspect
- 4 Hybrid Approach**
- 5 Conclusions

Temporal annotation

- It has matured recently
- Systems can be built to automatically annotate text with temporal information
- These systems can be used to expand the MRSs with further information about time
- Useful to add information that is difficult to process by the grammar
 - Describe time intervals more precisely
 - Check the consistency of the temporal relations included in the MRSs
 - Correct temporal relations on the basis of extralinguistic criteria, e.g. pragmatics

Temporal Annotation

- TimeML
 - Temporal expressions: `<TIMEX3 tid="t15" value="1998-02-27">Friday</TIMEX3>`
 - The document's creation time (our speech time)
 - Event terms: `<EVENT eid="e6">gave</EVENT>`
 - Temporal relations: `<TLINK eventID="e6" relType="BEFORE" relatedToTime="t15"/>`

Example

The mayor of Moscow has `<EVENT eid="e1">allocated</EVENT>` funds to `<EVENT eid="e2">help</EVENT>` `<EVENT eid="e3">build</EVENT>` a museum in honor of Mikhail Kalashnikov, the Russian who `<EVENT eid="e6">gave</EVENT>` his name to the world's most widely wielded weapon, `<EVENT eid="e91">according</EVENT>` to a news agency `<EVENT eid="e55">report</EVENT>` `<TIMEX3 tid="t15" value="1998-02-27">Friday</TIMEX3>`.

`<TLINK eventID="e6" relType="BEFORE" relatedToTime="t15"/>`

`<TLINK eventID="e91" relType="OVERLAP" relatedToTime="t15"/>`

`<TLINK eventID="e55" relType="OVERLAP" relatedToTime="t15"/>`

Temporal Annotation

- TimeML
 - Temporal expressions: `<TIMEX3 tid="t15" value="1998-02-27">Friday</TIMEX3>`
 - The document's creation time (our speech time)
 - Event terms: `<EVENT eid="e6">gave</EVENT>`
 - Temporal relations: `<TLINK eventID="e6" relType="BEFORE" relatedToTime="t15"/>`
- Automatic TimeML annotation tools
 - Mostly machine learning
 - LX-TimeAnalyzer: TimeML annotation of Portuguese [Costa and Branco 2012b, Costa and Branco 2012a]

Postprocess the MRSs output by LXGram

- Correct the temporal relations if necessary
- Add the normalized representation of the speech time
- Add normalized representations of other dates and times

Add the normalized representation of the speech time

- A feature T-VALUE is appropriate for temporal indices
- It holds the normalized value of the time interval that the index represents
- In the MRSs produced by LXGram it is left underspecified, or filled in with the value *speech-time*
- Postprocessing replaces this value with the normalized value of the corresponding <TIMEX₃>

Example

Choveu "It rained"

```
h1, e2
{ h3: at_rel(e2, t4)
  h3: is-before(t4, t5 { speech-time 2012-07-03T12:00:00 })
  h3: aspectual-operator_rel(e2, e6, h7)
  h7: _chover_v_rel(e6)
  { h1 =q h3 }
```

"rain" →

Add normalized representations of other dates and times

- In some cases, additional temporal relations between times and events are added

Example

Choveu hoje "It rained today"

h1, e2

{ h3: at_rel(e2, t4)

h3: is-before(t4, t5 { ~~speech-time~~ 2012-07-03T12:00:00 })

h3: aspectual-operator_rel(e2, e6, h7)

"rain" → h7: _chover_v_rel(e6)

"today" → h7: _hoje_a_rel(e9, e6, t8 { 2012-07-03 })

→ h7: include-or-equal(t8, t4) }

{ h1 =q h3 }

Open questions

- It might seem sensible to remove EPs for words in time expressions
- But it would cause problems
 - Time expressions may contain non-temporal content, which would be lost: *in that **cold** night ...*
 - Time expressions as arguments of verbs: *that day marked the beginning ...*

Example

Choveu hoje "It rained today"

h1, e2

{ h3: at_rel(e2, t4)

h3: is-before(t4, t5 { ~~speech-time~~ 2012-07-03T12:00:00 })

h3: aspectual-operator_rel(e2, e6, h7)

"rain" → h7: _chover_v_rel(e6)

"today" → h7: include-or-equal(t8 { 2012-07-03 }, t4) }

{ h1 =q h3 }

- ① Introduction
- ② Tense
- ③ Aspect
- ④ Hybrid Approach
- ⑤ Conclusions**

Summary

- HPSG implementation of tense and aspect
- It combines much of what is said in the literature
- A lot is left underspecified
- But it makes some of the temporal and aspectual meaning of sentences explicit
- Integration with an external component can fill in some of the missing information

- Bonami, O. 2002. A syntax-semantics interface for tense and aspect in French. In Eynde, F. V., Hellan, L., and Beermann, D., editors, *The Proceedings of the 8th International Conference on Head-Driven Phrase Structure Grammar*, pages 31–50, Stanford. CSLI Publications.
- Comrie, B. 1985. *Tense*. University Press, Cambridge.
- Costa, F. and Branco, A. 2012a. Extracting temporal information from Portuguese texts. In Caseli, H. et al., eds., *Computational Processing of the Portuguese Language—10th International Conference, PROPOR 2012*, volume 7243 of *Lecture Notes in Artificial Intelligence*, pages 99–105.
- Costa, F. and Branco, A. 2012b. LX-TimeAnalyzer: A temporal information processing system for Portuguese. Technical report, Universidade de Lisboa, Faculdade de Ciências, Departamento de Informática.
- Davidson, D. 1967. The logical form of action sentences. In Rescher, N., editor, *The Logic of Decision and Action*. University of Pittsburgh Press.
- de Swart, H. 1998. Aspect shift and coercion. *Natural Language and Linguistic Theory*, 16:347–385.
- de Swart, H. 2000. Tense, aspect and coercion in a cross-linguistic perspective. In Butt, M. and King, T. H., editors, *Proceedings of the Berkeley Formal Grammar conference*, Stanford. CSLI Publications.

- Dowty, D. R. 1979. *Word Meaning and Montague Grammar: the Semantics of Verbs and Times in Generative Semantics and Montague's PTQ*. Reidel, Dordrecht.
- Flouraki, M. 2006. Constraining aspectual composition. In Müller, S., editor, *The Proceedings of the 13th International Conference on Head-Driven Phrase Structure Grammar*, pages 140–157, Stanford. CSLI Publications.
- Goss-Grubbs, D. 2005. An approach to tense and aspect in Minimal Recursion Semantics. Master's thesis, University of Washington, Seattle, Washington.
- Kamp, H. and Reyle, U. 1993. *From Discourse to Logic: An Introduction to Modeltheoretic Semantics, Formal Logic and Discourse Representation Theory*. Kluwer, Dordrecht.
- Partee, B. 1973. Some structural analogies between tenses and pronouns in English. *The Journal of Philosophy*, 70:601–609.
- Pustejovsky, J. 1991. The syntax of event structure. *Cognition*, 41:47–81.
- Reichenbach, H. 1947. *Elements of Symbolic Logic*. University of California Press, Berkeley.
- Vendler, Z. 1967. Verbs and times. *Linguistics in Philosophy*, pages 97–12