LXGram in the Shared Task "Comparing Semantic Representations" of STEP 2008

António Branco
Francisco Costa
Universidade de Lisboa (Portugal)
email: Antonio.Branco@di.fc.ul.pt

Abstract

LXGram is a hand-built Portuguese computational grammar based on HPSG (syntax) and MRS (semantics). The LXGram system participated in the STEP 2008 shared task which aims at comparing semantic representations produced by NLP systems such as LXGram. Every participating team had to contribute a small text. The text that we submitted for the shared task was originally in Portuguese (an excerpt from a newspaper) and translated into English, to make a meaningful comparison at the shared task possible. Likewise, the English texts contributed by the other participating teams were translated into Portuguese. Because the LXGram generates many different analyses (mainly due to PP attachment ambiguities), the preferred analysis was selected manually. It was required to extend LXGram's lexicon and inventory of syntax rules to be able to get a reasonable performance on the shared task data. Eventually, our system was able to produce an analysis for 20 out of the 30 sentences of the shared task data.

1 Introduction

This paper describes the participation of the Portuguese grammar LXGram in the Shared Task of STEP 2008 "Comparing Semantic Representations" (Bos, 2008). This Shared Task was held in the University of Venice on 22–24 September 2008, with the purpose of comparing semantic representations produced by different natural language processing systems. This task had seven participating teams. Each team contributed with a small text (up to five sentences long) to be processed by all the systems.

LXGram is a hand-built, general purpose computational grammar for the deep linguistic processing of Portuguese. It is developed under the grammatical framework of Head-Driven Phrase Structure Grammar, HPSG (Pollard and Sag, 1987, 1994; Sag et al., 2003) and uses Minimal Recursion Semantics, MRS (Copestake et al., 2005) for the representation of meaning. This grammar implementation is undertaken with the LKB (Copestake, 2002) grammar development environment and its evaluation and regression testing is done via [incr tsdb()] (Oepen, 2001). It is also intended to be compatible with the PET parser (Callmeier, 2000).

The LinGO Grammar Matrix (version 0.9), an open-source kit for the rapid development of grammars based on HPSG and MRS, was used as the initial code upon which to build LXGram. The grammar is implemented in the LKB using the $\mathcal{TDL}$ formalism (Krieger and Schäfer, 1994), based on unification and on typed feature structures, and whose types are organized in a multiple inheritance hierarchy.

For more information, please refer to a detailed implementation report (Branco and Costa, 2008a) or on pages 31–43 of this volume (Branco and Costa, 2008b). A free version of the grammar can also be obtained at `http://nlx.di.fc.ul.pt/lxgram`, under an ELDA research license.

Section 2 introduces the main features of the Minimal Recursion Semantics format, which is employed in the semantic representations produced by LXGram. In Section 3, the sample text that the LXGram team submitted is described, together with an explanation of the representations derived by the grammar. Finally, Section 4 discusses the results for the full data set of the Shared Task.

2 Semantic Formalism

In LXGram, semantic information is encoded following Minimal Recursion Semantics (MRS) format for semantic representation (Copestake et al., 2005). MRS has several properties that makes it an interesting semantic representation format from the point of view of computational semantics.

Notoriously, it allows underspecification of the scope of relevant operators, which permits that a sentence with scope ambiguities can be given a single, underspecified representation. For some applications, for instance machine translation between closely related languages from the same language family, the underspecified representations may be sufficient and bring the benefit of avoiding possible combinatorial explosion into as many parses as readings.

In a nutshell, the underspecification of scope is achieved by associating every basic relation to a handle (in the feature structure for a relation, the feature LBL encodes this handle) and describing the constraints that hold between these handles (in the feature HCONS, handle constraints). These constraints can be stated in a way such that

some scope resolution options are allowed while others are discarded. Nevertheless, there may applications for which it may be important to get fully specified semantic representations. In this case, MRS permits that the different scope possibilities be computed on demand from the underspecified representation.

Also worth referring in this very brief presentation of the gist of MRS, it is the representation of conjunction with the relative order of conjuncts underspecified, by giving the same handle to the different conjuncts. This avoids computing associativity and commutativity of conjunction in situations where spurious overgeneration may arise.

Please consult Branco and Costa (2008a) in this volume (pages 31–43) for an example illustrating quantifier scope ambiguities and underspecification. Due to space limitations, it is not possible to provide further details on the MRS formalism here. For the presentation of MRS, please consult Copestake et al. (2005).

3 Sample Text

The following sentences are our examples for the shared task:

(1) A primeira escola de treino de cães-guias do País vai
the first school of training of leader dogs of the country goes
nascer em Mortágua e treinará 22 cães-guias por ano.
to be born in Mortágua and will train 22 leader dogs per year

The first school for the training of leader dogs in the country is going to be created in Mortágua and will train 22 leader dogs per year.

(2) Em Mortágua, João Pedro Fonseca e Marta Gomes coordenam o
in Mortágua João Pedro Fonseca and Marta Gomes coordinate the
projecto que sete pessoas desenvolvem nesta escola.
project that seven people develop in this school

In Mortágua, João Pedro Fonseca and Marta Gomes coordinate the project that seven people develop in this school.

(3) Visitaram vários espaços semelhantes em Inglaterra e em França,
they visited several spaces similar in England and in France,
e numa das escolas francesas estão já em estágio duas
and in one of the schools French are already in internship two
futuras treinadoras.
future trainers

They visited several similar places in England and in France, and two future trainers are already doing internship in one of the French schools.

(4) Os fundos comunitários asseguram a manutenção da escola até
the funding communitarian ensure the maintenance of the school until
1999.
1999

The communitarian funding ensures the operation of the school until 1999.

(5) Gostaríamos que a nossa escola funcionasse à semelhança das
we would like that the our school worked to the similarity of the
francesas, que vivem de dádivas, do merchandising e até
French which live from donations from the merchandising and even
das rifas que as crianças vendem nas escolas.
from the raffles that the children sell in the schools

We would like our school to work similarly to the French ones, which live from donations, from the merchandising and even from the raffles that children sell in school.

These sentences were adapted from newspaper text. We have chosen them because they display interesting phenomena.

The semantic representations that LXGram produces for these sentences are presented at the Shared Task website `http://www.sigsem.org`. An example is included in Appendix B. Several analyses are obtained for these examples (e.g. one of the sentences got 540 parses), the main reason being PP attachment ambiguity. The semantic representations we present are the ones associated to the preferred analyses, which were selected manually.

Note that since the representations could not be displayed in a single page, the value of the feature RELS was split across multiple pages. To ensure readability, the values of the other features (LTOP, INDEX and HCONS) are repeated on every page pertaining to the same representation.

Some comments are in order concerning these representations:

- The morphological person, number and gender are encoded as features (PERSON, NUMBER, GENDER) of the relevant index (quantified variable) that is present there. For indices, the boolean feature DIV is also used, that shows the value + for plurals and mass nouns.

- Event variables are included for the relations introduced by verbs, adjectives, prepositions and adverbs (under their ARG0 feature). The morphological information on the verbs is also encoded as features of these events. This is the purpose of the features MOOD, TENSE and ASPECT. There is also a feature SF (sentence force) that represents whether a sentence denotes a proposition, a question or a command. The feature ELLIPTICAL-PUNCT denotes whether the sentence ends with an ellipsis (...) and is useful in order to constrain what is generated by the grammar.

- There is a *tense_rel* relation associated to each verb form. Its ARG0 feature is the same as the ARG0 of the verb it is associated with. The purpose of this extra relation is to make an event variable present in the semantic representations for the copular sentences where the relevant predicate is provided by a noun (none of these examples). In such cases this event will contain the morphological information of the copular verb.

- Note that the information about whether adjectives have intersective semantics (see "francês"—"French"—in sentence (3)) or non-intersective semantics (see

"futuro"—"future"—in sentence (5)) is visible in the corresponding semantic representations.

The names of the predicates that correspond to lexical items of several classes (common nouns, verbs, adjectives, adverbs, prepositions, etc.) follow a naming convention that includes a lemma field, a part-of-speech field and an optional sense field (often reflecting subcategorization). Table 1 lists the predicates present in these representations and provides the corresponding English lemmas. There are other special relations in these representations:

- *udef_q_rel*
  the quantifier for bare NPs
- *proper_q_rel*
  the quantifier for proper names
- *tense_rel*
  associated to every verbal relation (see discussion above)
- *named_rel*
  associated to proper names
- *name-precedes_rel*
  associated to proper names
- *string-equals_rel*
  equality between strings
- *indef_q_rel*
  associated to some indefinites. In particular it is the quantifier used for NPs that are introduced by elements that can also follow determiners (e.g. cardinals and vague quantifiers like "vários"—"several")
- *cardinal_rel*
  constrains the cardinality of the set denoted by the expression linked to its ARG1 feature
- *greater-or-equal_rel*
  the integer in its ARG0 is greater than or equal to the integer in its ARG1 feature
- *plus_rel*
  the integer in its ARG0 is the result of summing the two integers in the TERM0 and TERM1 features
- *int-equals_rel*
  equality between integers
- *ellipsis-or-generic_n_1_rel*
  placeholder relation when there are missing nouns

Table 1: Correspondence of Portuguese MRS relations and English lemmas

MRS Relation	**English lemma**
_ano_n_rel	year
_à_semelhança_a_-de-_rel	similarly
_assegurar_v_rel	to ensure
_até_a_rel	even
_até_p_rel	until
_cão-guia_n_rel	leader dog
_comunitário_a_rel	communitarian
_coordenar_v_rel	to coordinate
_criança_n_rel	child
_dádiva_n_-de-a-_rel	donation
_de_p_rel	of, from
_desenvolver_v_rel	to develop
_e_coord_rel	and
_em_p_rel	in
_espaço_n_rel	space
_estágio_n_rel	internship
_este_a_rel	this
_escola_n_rel	school
_francês_a_rel	French
_funcionar_v_rel	to work
_fundo_n_rel	funding
_futuro_a_rel	future
_gostar_v_rel	to like
_ir_v_aux_rel	to be going to
_já_a_rel	already
_manutenção_n_-de-por-_rel	maintenance
_merchandising_n_rel	merchandising
_nascer_v_rel	to be born
_o_q_rel	the
_país_n_rel	country
_por_p_rel	per
_pessoa_n_rel	person
_primeiro_a_rel	first
_projecto_n_-de-por_rel	project
_rifa_n_rel	raffle
_semelhante_a_-a-_rel	similar
_treinador_n_-de-_rel	trainer
_treinar_v_rel	to train
_treino_n_-de-por-_rel	training
_um_q_rel	a
_vários_a_scop_rel	several
_vender_v_-a-_rel	to sell
_visitar_v_rel	to visit
_viver_v_rel	to live

Sometimes some details of the semantic representations that are possible to obtain depend on the features of the system where LXGram is developed and runs. In particular, for each feature that represents an argument of a relation (ARG0, ARG1, ARG2, CARG, ...), it must be stated in the configuration files whether it will contain a constant (e.g. a string literal). For instance, we must say that the feature CARG always contains a value, for visualization purposes. This fact sometimes constrains the display of the semantic representations. It is the reason why the semantics for proper names and for cardinals is more copious than what would seem necessary at first.

For instance, the semantics associated to "7 pessoas" ("7 people") in sentence (2) is roughly $\lambda x.cardinal\_rel(e, \_pessoa\_n\_rel(x), j_1) \wedge greater\text{-}or\text{-}equal\_rel(j_1, j_2) \wedge int\text{-}equals(j_2, 7)$ (note that conjunction is denoted in MRS via identical labels for relations). The information conveyed by the last two predicates could be simply given by $greater\text{-}or\text{-}equal\_rel(j_1, 7)$. However, for that to display correctly we would have to configure the system to display the second argument of the *greater-or-equal_rel* relation as a constant. This will not always be the case: in the semantics for "22" that argument is the integer that is the result of summing "20" and "2" (number expressions receive compositional semantics), represented with the help of the *plus_rel* relation. The LKB does not allow one to compute arithmetic expressions.

These few sentences present some interesting problems for the computation of semantic representation in general.

Typically, one is not able to resolve missing nouns, as this sometimes requires access to pragmatic information. As a consequence, the semantics produced for sentences with a missing noun (see sentence (5)) includes an *ellipsis-or-generic_n_1_rel* instead of the relation corresponding to that noun.

Also, it is very hard if not impossible to recover missing arguments. See for instance the semantics for the adjective "semelhante" ("similar") in sentence (3). The missing argument is given the type *r*, instead of the type *x* of quantified variables, so that we can omit a quantifier for it in the semantics and still be able to ask the system for scoped solutions (the system would complain about free variables if these elements were given the type *x*).

Finally, it is worth noting that there are some limitations of the semantic representations obtained given that the empirical coverage of the grammar is still in development. Currently, the grammar does not make yet any distinction between restrictive and non-restrictive relative clauses, as we have not focused on the fully-fledged implementation of the semantics of non-restrictive relative clauses yet. This can be seen in the semantics for the last example, where both relative clauses are semantically combined with their head in the same way.

4 Performance in the Shared Task

There are seven small texts in the Shared Task. The sample text we submitted is text 4. We translated the other six texts into Portuguese before passing them to the system.

Translation of the Texts

The translations were done by the authors. We tried to make them as literal as possible in order to support comparability of the different systems taking part in the Shared Task, but some bits were not literally translated as that would have produced unnatural

sentences. We also tried not to make the texts easy to parse by the system by simplifying the texts in the translations. We present the translation for the texts 1, 2, 3, 5, 6 and 7 in the Appendix A, with English glosses.

Initial Coverage

When we tried to parse the other six texts of the Shared Task, we got 0% coverage. The causes for parse failure were missing words in the lexicon and missing syntactic constructions.

Since the aim of the Shared Task is not to evaluate data coverage but rather to compare the semantic representations output by different NLP systems, we made an effort to expand LXGram by enlarging the lexicon and implementing some syntax rules, with the purpose of producing semantic representations for as many sentences in the Shared Task data as possible, within the time constraints.

During this grammar expansion, we tried not to tune the grammar to these particular sentences. We tried to make the implementation of new phenomena general. For this reason, some phenomena were not implemented deliberately, because we felt that we would not be able to produce general solutions for them within the time limit. This is the case of WH- questions (present in the first text), which are not yet supported by LXGram and whose implementation we did not want to rush.

Grammar Expansion

We added 97 lexical entries to the grammar. For some of these items, we had to create new lexical types, because they have subcategorization frames for which there was still no lexical type in the grammar. One example is the noun "pedido" (*order*), which was implemented as having two arguments realized by prepositional phrases, the first one headed "de" and the second one headed by "a". LXGram already contained lexical types for nouns with two arguments, but introduced by different prepositions. Although these two arguments of the noun were not present in the example where this noun occurs (the third sentence of text 3), we nevertheless created a new lexical type for this subcategorization frame. We could have used an existing lexical type for nouns with no complements and that particular sentence would have parsed fine, but the predicate for that noun would not be a two-place predicate in the MRS representation. We added 10 new lexical types.

The constructions that were implemented in LXGram in order to parse these sentences were:

- the progressive. In European Portuguese, the progressive is expressed via a form of the verb "estar" (*to be*) combined with an infinitive preceded by the preposition "a".

- temporal expressions headed by the verb "haver" (*there to be*). The temporal expression *for some time* (second sentence of text 2) is expressed in Portuguese as "há algum tempo" (literally: *there is some time*). The verb form cannot be analyzed as a preposition, because this sort of expression is syntactically compositional. For instance, the verb inflects for tense (it can appear in the imperfect if the main verb of the clause is in a past tense) and there can be adverbs modifying it to its right ("há já algum tempo", *there is already some time*, i.e.

for some time now). We created a unary syntax rule that takes as daughter a clause headed by this verb and produces a mother node with the syntactic characteristics of a clause introduced by a subordinating conjunction and modifying another clause. This rule adds a relation similar to a relation introduced by a subordinating conjunction, and it's called *abstract-temporal_x_rel*. We take this relation as having the meaning of "since", but with the two arguments reversed, and the Portuguese clause for *that is known for some time* gets analyzed as meaning roughly *there is some time (some time has passed) since that is known*. That is a very literal semantic representation, but it allows us to keep the semantic composition mechanism completely monotonic.

- the impersonal pronoun "se". The most naturally sounding translation of *it was suspected that* (last sentence of text 5) is "suspeitou-se que", with a verb in the active voice and its subject being realized by a clitic pronoun. This clitic has to appear adjacently to the verb, which is atypical for subjects in Portuguese.
- NP appositives. We also implemented a rule to allow NP apposition. This was because of sentences like the second sentence in text 6.

Additionally, a few preprocessor rules were expanded. For instance, sentences like the last sentence of text 7 require integer literals to be considered as proper names. We cannot create lexical entries for all integers, so we added preprocessor rules in order to contemplate the possibility of integers as proper names.

Final Results

After grammar expansion, 20 sentences out of the 30 sentences in all the texts of the Shared Task got an analysis. The sentences that could not be parsed are the following:

- Text 1: sentences (c) and (d).
- Text 5: sentences (a), (c) and (d)
- Text 6: all sentences
- Text 7: sentences (a) and (b)

The two sentences of text 1 that could not be parsed contain WH- questions, which are currently not supported by the system.

The sentence (a) of text 5 could not be parsed because it contains two sentences as the complement of a verb. LXGram cannot yet combine two independent sentences, and we chose to not implement this possibility because the combination of an n-way ambiguous sentence with another m-way ambiguous sentence would be n $\times$ m-way ambiguous.

The sentence (c) of the same text was not parsed because of a semantically vacuous clitic (not implemented yet) and a relative clause modifying another clause (also not covered). LXGram does not support sentence relatives and we chose not to implement them yet because, if the relative pronoun is filling a subject position (as in that sentence), the verb has to allow for propositional subjects. In LXGram, we currently only have subcategorization frames for verbs that take NPs as subjects, and we have to review all lexical entries for verbs before we can parse that sentence.

For the remaining sentences without a parse, the reason was efficiency. Several of the sentences in the Shared Task data translate to Portuguese sentences that are very long (over 40 words) or have a very high number of prepositions, producing many attachment possibilities. Note that we were doing exhaustive search. In many cases the parser would run out of memory. In order to alleviate this problem, we used the PET parser instead of the LKB parser for the longer sentences. PET is considerably faster, because it is implemented in C (the LKB is in Lisp), and it precompiles the grammar into a binary format. Also, the input to PET can be preprocessed by a POS tagger, in order to reduce lexical ambiguity. We did this preprocessing for some of the longer sentences.

However, PET dumps MRS representations as text, and choosing the best parse from this sort of output is not practical, especially for sentences with many readings. So we exported the results into a format that can be read by [incr tsdb()], a tool for the management of test suites and corpora. With this tool, it is possible to choose parses by choosing discriminants derived from all analyses. Choosing or rejecting a single discriminant can eliminate a large number of analyses in one step. However, [incr tsdb()] calls the LKB to reconstruct the trees based on the output of PET (which includes the names of the rules used and syntactic constituency), when one wants to choose the best parse. Even though the parse forest has already been built by the PET parser, the LKB can still run out of memory when it is reconstructing the feature structures if the number of analyses is sufficiently large (we had a sentence with over 18000 parses).

We also tried commenting out some rules that were not necessary to parse these sentences, with the purpose of reducing the search space. Examples include robustness rules, for parsing strings with no verb.

In the near future, we will be working on a stochastic disambiguation module, which PET supports, in order to constrain the parser's search space and to keep only the best *n* parses, so that we can avoid the efficiency problems that we are facing at the moment.

Analyses

The semantic representations for the sentences that LXGram parsed successfully are presented in the appendix. As mentioned before, we performed exhaustive search. We chose the best parse manually.

We used [incr tsdb()] associated to the LKB in order to choose the preferred reading. After that we exported the MRS representation. The LKB exports LaTeX directly. We edited the exported LaTeX in order to make the representations fit into the pages of the appendix. This involved manually adding newlines and page breaks. We also corrected characters with diacritics, which did not display correctly, and we removed characterization information: after the name of each predicate, there is a pair of character positions indicating the substring in the input spanned by the lexical items or rules associated to that predicate; they were removed because they are not interpretable by someone who does not know the implementation details, e.g. the semantics for null subjects span the substring of the entire VP since this piece of semantics is introduced by a unary rule that takes a VP as daughter.

Discussion of the Results

We would like to comment on some of the semantic representations obtained with LXGram.

As we have pointed out before, some details of the semantics are not completely independent of language. For an example, see the discussion above about temporal expressions headed by the verb "haver".

MRS does not directly support a treatment of intentionality. For instance, sentence (c) of text 2 contains an intentional context: it does not assert the existence of "other cancers caused by viruses". There is no standard way of representing this sort of intentionality with MRS.

Also, MRS does not support conjunction of quantifiers. There is no MRS equivalent to a lambda expression like $\lambda P.Quant_1(x,P(x)) \wedge Quant_2(y,P(y))$. The usual MRS representations associated with NP coordination have to include an explicit relation for the truth function involved (but taking referential indices as arguments), as well as an extra quantifier relation (the relation used in these cases is called *udef_q_rel*, which is also the name for the quantifier of bare NPs).

Some phenomena are difficult to analyze. An example is in sentence (c) of text 7. In the Portuguese translation, we have two coordinated NPs at the end of the sentence (the best sounding translation requires a determiner before each of the two nouns), which are followed by a PP. The Portuguese translation interprets this PP as realizing an argument of both nouns (cf. *federal government interest and federal government tax incentives*). We could not get this reading, because we do not allow PP arguments to attach higher than determiners. The analysis that we present leaves the first noun with this argument underspecified, as this PP attaches directly to the second noun in the corresponding syntax tree. This possibility of PP attachment seems to be required for cases of NP coordination like this one, but it can be a source of overgeneration for NPs that are not coordinated. This phenomenon affects other NP elements, like adjective phrases, that can also take scope over a coordination of NPs. The current implementation forces all noun dependents that have a restrictive interpretation to attach lower than determiners, as that is the place where the restrictor of the quantifier for that NP is visible in the feature structures.

References

Bos, J. (2008). Introduction to the Shared Task on Comparing Semantic Representations. In J. Bos and R. Delmonte (Eds.), *Semantics in Text Processing. STEP 2008 Conference Proceedings*, Volume 1 of *Research in Computational Semantics*, pp. 257–261. College Publications.

Branco, A. and F. Costa (2008a). A computational grammar for deep linguistic processing of Portuguese: LXGram, version A.4.1. Technical report, University of Lisbon, Department of Informatics.

Branco, A. and F. Costa (2008b). High Precision Analysis of NPs with a Deep Processing Grammar. In J. Bos and R. Delmonte (Eds.), *Semantics in Text Processing. STEP 2008 Conference Proceedings*, Volume 1 of *Research in Computational Semantics*, pp. 31–43. College Publications.

Callmeier, U. (2000). PET — A platform for experimentation with efficient HPSG processing techniques. *Natural Language Engineering 6*(1), 99–108. (Special Issue on Efficient Processing with HPSG).

Copestake, A. (2002). *Implementing Typed Feature Structure Grammars*. Stanford: CSLI Publications.

Copestake, A., D. Flickinger, I. A. Sag, and C. Pollard (2005). Minimal Recursion Semantics: An introduction. *Journal of Research on Language and Computation 3*(2–3), 281–332.

Krieger, H.-U. and U. Schäfer (1994). $\mathcal{TDL}$ — A type description language for constraint-based grammars. In *Proceedings of the 15th International Conference on Computational Linguistics*, Kyoto, Japan, pp. 893–899.

Oepen, S. (2001). [incr tsdb()] — competence and performance laboratory. User manual. Technical report, Computational Linguistics, Saarland University, Saarbrücken, Germany. In preparation.

Pollard, C. and I. Sag (1987). *Information-Based Syntax and Semantics, Vol. 1*. Number 13 in CSLI Lecture Notes. Stanford: CSLI Publications.

Pollard, C. and I. Sag (1994). *Head-Driven Phrase Structure Grammar*. Stanford: Chicago University Press and CSLI Publications.

Sag, I. A., T. Wasow, and E. M. Bender (2003). *Syntactic Theory – A Formal Introduction* (2nd ed.). Stanford: CSLI Publications.

Appendix A: Translations of the Texts for the Shared Task

Text 1

(1) Um objecto é lançado com uma velocidade horizontal de 20 m/s de um penhasco que tem 125 m de altura.
an object is thrown with a speed horizontal of 20 m/s from a cliff that has 125 m of height

An object is thrown with a horizontal speed of 20 m/s from a cliff that is 125 m high.

(2) O objecto cai pela altura do penhasco.
the object falls for the height of the cliff

The object falls for the height of the cliff.

(3) Se a resistência do ar é negligenciável, quanto tempo demora o objecto a cair ao chão?
if the resistance of the air is negligible how much time takes the object to fall to the ground

If air resistance is negligible, how long does it take the object to fall to the ground?

(4) Qual é a duração da queda?
what is the duration of the fall

What is the duration of the fall?

Text 2

(1) O cancro cervical é causado por um vírus.
the cancer cervical is caused by a virus

Cervical cancer is caused by a virus.

(2) Isso é conhecido há algum tempo e levou a uma vacina que parece preveni-lo.
that is known there is some time and led to a vaccine that seems to prevent it

That has been known for some time and it has led to a vaccine that seems to prevent it.

(3) Os investigadores têm procurado outros cancros que possam ser causados por vírus.
the researchers have looked other cancers that may be caused by viruses

Researchers have been looking for other cancers that may be caused by viruses.

Text 3

(1) O John foi a um restaurante.
the John went to a restaurant

John went into a restaurant.

(2) Havia uma mesa no canto.
there was a table in the corner

There was a table in the corner.

(3) O empregado anotou o pedido.
the waiter wrote down the order

The waiter took the order.

(4) A atmosfera era acolhedora e simpática.
the atmosphere was warm and friendly

The atmosphere was warm and friendly.

(5) Ele começou a ler o seu livro.
he began to read the his book

He began to read his book.

Text 5

(1) Enquanto os 3 canhões do torreão 2 eram carregados, um membro da equipa que estava a
as the 3 guns of the Turret 2 were loaded a member of the crew who was to
operar o canhÃčo central gritou ao telefone "Tenho aqui um problema. Ainda não estou
operate the gun central yelled to the phone I have here a problem. Still not I am
preparado".
ready

As the 3 guns of Turret 2 were being loaded, a crewman who was operating the center gun yelled into the phone, "I have a problem here. I am not ready yet."

(2) Então o explosivo rebentou.
then the propellant exploded

Then the propellant exploded.

(3) Quando os membros da equipa do canhão morreram, estavam agachados de forma não
when the members of the crew of the gun died they were crouching of way not
natural, o que sugeria que sabiam que se daria uma explosão.
natural which suggested that they knew that DUMMY CLITIC would happen an explosion

When the gun crew was killed they were crouching unnaturally, which suggested that they knew that an explosion would happen.

(4) O explosivo que foi usado era feito de pedaços de nitrocelulose que foram produzidos
the propellant that was used was made from chunks of nitrocellulose that were produced
durante a Segunda Guerra Mundial e foram reembalados em 1987 em sacos que foram feitos
during the second world war and were repackaged in 1987 in bags that were made
em 1945.
in 1945

The propellant that was used was made from nitrocellulose chunks that were produced during World War II and were repackaged in 1987 in bags that were made in 1945.

(5) Inicialmente, suspeitou-se que este armazenamento poderia ter
initially suspected IMPERSONAL SUBJECT that this storage might have
reduzido a estabilidade da pólvora.
reduced the stability of the powder

Initially it was suspected that this storage might have reduced the powder's stability.

Text 6

(1) Entre as filas cerradas de casas do norte de Filadélfia, uma quinta urbana pioneira
amid the rows tightly packed of houses of the north of Philadelphia a farm urban pioneering
está a produzir comida local fresca para uma comunidade que frequentemente não a tem, e a
is to produce food local fresh for a community that often not it has and to
gerar dinheiro com isso.
generate money with it

Amid the tightly packed row houses of North Philadelphia, a pioneering urban farm is providing fresh local food for a community that often lacks it, and making money in the process.

(2) Greensgrow, um terreno de um acre de canteiros elevados e estufas no local de uma
Greensgrow a plot of one acre of beds raised and greenhouses on the site of a
antiga fábrica de galvanização de aço, está a ter lucro vendendo os próprios vegetais e
former factory of galvanization of steel is to have profit selling the own vegetables and
ervas assim como uma gama de produtos de agricultores locais, e gerindo um viveiro que
herbs as well as a range of products from farmers local and managing a nursery that
vende plantas e plântulas.
sells plants and seedlings

Greensgrow, a one-acre plot of raised beds and greenhouses on the site of a former steel-galvanizing factory, is turning a profit by selling its own vegetables and herbs as well as a range of produce from local growers, and by running a nursery selling plants and seedlings.

(3) A quinta lucrou cerca de 10000 dólares com uma receita de 450000 dólares em 2007, e
the farm earned about 10000 dollars with a revenue of 450000 dollars in 2007 and
espera ter um lucro de 5% sobre os 650000 dólares de receitas neste ano, o seu 10Âž ano,
hopes to have a profit of 5% on the 650000 dollars of revenue in this year the its 10th year
para poder abrir outra actividade noutro sítio de Filadélfia.
in order to be able to open another operation in another place of Philadelphia

The farm earned about $10,000 on revenue of $450,000 in 2007, and hopes to make a profit of 5 percent on $650,000 in revenue in this, its 10th year, so it can open another operation elsewhere in Philadelphia.

Text 7

(1) O desenvolvimento moderno da tecnologia e aplicações de energia eólica
the development modern of the techonology and applications of energy wind.ADJECTIVE
já estava numa fase avançada nos anos 30, quando por estimativa cerca de 600000
already was in a phase advanced by the years 30 when by estimation about 600000
moinhos forneciam áreas rurais com electricidade e serviços de bombeamento de água.
mills supplied areas rural with electricity and services of pumping of water

Modern development of wind-energy technology and applications was well underway by the 1930s, when an estimated 600,000 windmills supplied rural areas with electricity and water-pumping services.

(2) Quando a distribuição em larga escala de electricidade chegou às quintas e às terras
when the distribution in broad scale of electricity arrived to the farms and to the small
pequenas, o uso de energia eólica nos Estados Unidos começou a diminuir, mas
towns the use of energy wind.ADJECTIVE in the United States started to subside but
voltou a subir depois da falta de petróleo nos EUA no começo dos anos
it went back to raise after of the shortage of oil in the US in the beginning of the years
70.
70

Once broad-scale electricity distribution spread to farms and country towns, use of wind energy in the United States started to subside, but it picked up again after the U.S. oil shortage in the early 1970s.

(3) Nos últimos 30 anos, a investigação e o desenvolvimento têm oscilado de acordo
in the last 30 years the research and the development have fluctuated in accordance
com o interesse e os benefícios fiscais do governo federal.
with the interest and the benefits fiscal of the government federal

Over the past 30 years, research and development has fluctuated with federal government interest and tax incentives.

(4) Em meados dos anos 80, as turbinas eólicas tinham tipicamente uma potência
in middle of the years 80 the turbines wind.ADJECTIVE had typically a power rating
máxima de 150 kW.
maximum of 150 kW

In the mid-'80s, wind turbines had a typical maximum power rating of 150 kW.

(5) Em 2006, as turbinas comerciais de grande escala são comummente avaliadas em mais de 1
In 2006 the turbines commercial of large scale are commonly rated at more than 1
MW e estão disponíveis em no máximo 4 MW de capacidade.
MW and are available in at the most 4 MW of capacity

In 2006, commercial, utility-scale turbines are commonly rated at over 1 MW and are available in up to 4 MW capacity.

Appendix B: MRS Representation for Text 4, Sentence 1