

QUESTION GENERATION GUIDELINES

TASK DESCRIPTION: WHAT IS NEW

DOCUMENT COLLECTIONS: There are now two collections, JRC-Acquis (as last year) and Europarl.

Exactly as last year, we are using the same subset of the JRC-Acquis¹ corpus about European legislation that has parallel translations aligned at paragraph level in Bulgarian, Dutch, English, French, German, Italian, Portuguese, Romanian and Spanish.

In addition, we have part of Europarl which is a collection of the Proceedings of the European Parliament dating back to 1996. It comprises translations of each of the 11 official languages of the European Union (Danish, German, Greek, English, Spanish, Finnish, French, Italian, Dutch, Portuguese and Swedish). With the enlargement of the European Union to 25 member countries in May 2004, the European Union has begun to translate texts into even more languages. Translations into Bulgarian and Romanian are present starting from January 2009.

A (very small) subset of the Europarl has been created with parallel documents in all the 9 languages involved in the track (Bulgarian, Dutch, English, French, German, Italian, Portuguese, Romanian and Spanish) by crawling the web to get the data from the European Parliament's website, starting from 01/01/2009. The sub-set consists of roughly 50 parallel and aligned documents per language. N.B.: In order to facilitate the identification of the paragraphs, each of them has been given a unique progressive ID number inside each document.

Both subsets of the collections can be obtained from <http://celct.isti.cnr.it/ResPubliQA/Downloads>

¹ <http://wt.jrc.it/lt/Acquis/>

QUESTIONS: The test set will include a pool of 200 questions only in the languages of the tasks which have been activated (i.e. tasks for which at least TWO have registered!).

ANSWERS: As discussed in the Track Guidelines, there are two tasks this year:

1. **PARAGRAPH SELECTION (PS) TASK:** to retrieve one paragraph (Text+ID) containing the answer to a question in natural language. This task is very similar to the one performed last year.

2. **ANSWER SELECTION (AS) TASK:** beyond retrieving a paragraph, systems are required to retrieve also the exact answer (shorter string of text) answering a question in natural language.

It follows therefore that for each question generated we need to specify the single paragraph which includes the answer (as last year) and in addition we need to demarcate the substring of that paragraph which comprises the exact answer.

LANGUAGES INVOLVED: parallel-aligned documents are available in 9 languages, i.e: Bulgarian, Dutch, English, French, German, Italian, Portuguese, Romanian and Spanish.

IMPORTANT DATES

Question Generation	from February 20 to March 20
Corrections	by April 2
Question selection	by April 9
Translations of questions	from April 12 to April 30
Proof-reading of the question sets	by May 9
Test set release	May 17
Run submissions	May 27
Assessments of runs	from May 31 to June 21
Results to the participants	June 25
Contributions for the Overview	July 10
CLEF Workshop	20-23 September 2010, Padua, Italy

Test Sets Release: May 17, 2010 (12.01 pm –noon- Central European Summer Time).

The participants will have 5 DAYS to upload their submissions, starting from the moment when the questions are downloaded.

Final Deadline for Run Submission: May 27th, 2010 (11:59 pm –midnight- Central European Summer Time).

PARTNERS INVOLVED IN THE TRACK ORGANIZATION

- **ES:** Universidad Nacional de Educación a Distancia, Spain (UNED) – (general Co-ordination)
- **IT:** Center for the Evaluation of Language and Communication Technologies, Trento, Italy (CELCT) – (General Co-ordination)
- **EN:** University of Limerick, Ireland (LIM) - (General Co-ordination)
- **RO:** UAIC and RACAI, Romania.

I. OUTLINE AND DATES

- **stage 1) *Creation of the Questions (by March 20)***
Each organizing group (**CELCT, LIM, UAIC, UNED**) produces **75 questions** in its own language, searches for answers in the ACQUIS or EuroParl parallel collections, and provides also the English translation of the questions.

For more details about the questions collection see section II.

- **stage 2) *Corrections (by April 4)***
The English translations of the 300 generated questions are checked by native speakers at Limerick University and corrections from each group will be requested and entered via the web-interface.
- **stage 3) *Question selection***
300 candidate questions are obtained in total, and 200 questions will be chosen by the organizers out of the 300 produced to be used in the test set.
- **stage 4) *Translations of questions (by April 30)***
The final pool of 200 questions is then translated **only into the languages of the tasks for which at least two groups have registered** using English as inter-lingua. Information about registrations to the cross-language sub-tasks will be available by March 30.
While doing the translation, each group should also make sure that a paragraph containing the answer to the question exists in its own language collection.
- **stage 5) *Proof-reading (by May 9)***
Each question set will be proof-read by each language group in order to check spelling and wording errors, before putting test sets online.

Like in the previous year, web-based interfaces will be provided to help in the creation and translation of the questions.

More details, together with the URL, will be provided at the beginning of the question generation phase.

N.B.: Deadlines must be considered as hard deadlines. Please try to meet them!

II. BUILDING THE QUESTION SET

Each organizing group filling in the online interface:

- ❖ produces 75 questions in its own language
- ❖ provides the English translation of the questions
- ❖ searches for the paragraph containing the answer in the respective ACQUIS or Europarl parallel collections
- ❖ copies the id of the paragraph (*p_id*) in its own source language collection
- ❖ copies the id of the document from which the paragraph has been extracted
- ❖ demarcates the exact answer
- ❖ searches for the paragraph containing the answer in the ACQUIS or EuroParl English parallel documents, also making sure that the answer is contained in one paragraph only.
- ❖ copies the id of the paragraph (*p_id*) in the English collection
- ❖ demarcates the exact English answer

See the section III for the description of the types of questions.

RECOMMENDATIONS

- Please pay attention at the way in which questions are expressed: try to avoid artificial questions that suggest back-formulation of statement in the original source document making questions appear somehow unnatural . Questions must be well-formed, as posed by a person that does not know the answer nor its position in the text. In principle, the word overlapping between the question and the target document should be kept at a minimum.
- Questions must be created from the <body> of the document – not from passages which have no unique identifier (e.g. <NOTE> in ACQUIS).

III. QUESTION TYPE DESCRIPTION

Questions must NOT be grouped in series and must NOT contain anaphoric links to other entities. Each question must be guaranteed to have at least one paragraph containing the answer in the document collection (NO NIL questions should be produced).

As far as the question types are concerned, the following question categories will be considered:

- a) **Factoid**
- b) **Definition**
- c) **Reason-Purpose**
- d) **Procedure**
- e) **Opinion**
- f) **Other**

No NIL questions must be produced, and in the case of LIST questions, all the requested items must be included in the response to be returned: a paragraph in the case of the PS task and a short string of text in the AS task.

- a) **Factoid questions** are fact-based questions, asking for the name of a person, a location, the extent of something, the day on which something happened, etc.

Examples:

Q: *What are the foodstuffs for particular nutritional use intended for?*

P: 2. This Directive covers foodstuffs for particular nutritional use fulfilling the particular requirements of infants and young children in good health in the Community and are intended for use by infants while they are being weaned, and by young children as a supplement to their diet and/or for their progressive adaptation to ordinary food. They comprise:

E: infants and young children in good health

Q: *In how many languages is the Official Journal of the Community published?*

P: The Official Journal of the Community shall be published in the four official languages.

E: four official languages

b) Definition questions are questions such as "What/Who is X?", questions asking for the role/job/important information about someone, questions asking for the mission/full name/important information about an organization

Examples:

Q: *What does IPP denote in the context of the environmental policies?*

P: Since then, new policy approaches on sustainable goods and services have been developed. These endeavours undertaken at all political levels have culminated in the Green Paper on Integrated Product Policy(1) (IPP). This document proposes a new strategy to strengthen and refocus product-related environmental policies and develop the market for greener products, which will also be one of the key innovative elements of the sixth environmental action programme - Environment 2010: "Our future, our choice"(2).

E: Integrated Product Policy(1)

c) Reason-purpose questions: questions asking for the reasons/goals for something. happening

Examples:

Q: *Why have the imports of live poultry from Romania been suspended?*

P: (2) Commission Decision 2005/710/EC of 13 October 2005 concerning certain protection measures in relation to highly pathogenic avian influenza in Romania [5] provides that Member States are to suspend imports of live poultry, ratites and farmed and wild feathered game and hatching eggs of those species from the whole territory of Romania and of certain products from birds from parts of that territory.

E: highly pathogenic avian influenza in Romania

Q: *What is the Commission's aim in establishing financial forecasts each year, broken down by category of expenditure, for the three subsequent financial years?*

P: In order to place the budget of the Communities within a framework of forward planning for several years, the Commission shall, each year, after receiving the Opinion of the Budgetary Policy Committee, draw up a financial forecast for the three subsequent financial years, showing the financial implications for the Community resulting from Regulations and Decisions in force and from proposals submitted by the Commission to the Council. The forecast shall be broken down by category of expenditure.

E: to place the budget of the Communities within a framework of forward planning for several years

d) Procedure questions: questions asking for a set of actions which is the official or accepted way of doing something:

Examples:

Q: *What is the procedure to import an animal to the Community from a non-member country?*

P: Animals imported from non-member countries must, with respect to the requirements for their welfare, whilst being reared, receive treatment at least equivalent to that guaranteed by this Directive to animals of Community origin. This requirement must be stated in a certificate issued by the competent authority of the non-member country concerned.

E: whilst being reared, receive treatment at least equivalent to that guaranteed by this Directive to animals of Community origin. This requirement must be stated in a certificate issued by the competent authority of the non-member country concerned.

Q: *How do you measure the stopping distance of a tractor?*

P: 1. BRAKING TESTS 1.1. General 1.1.1. The performance prescribed for service braking devices shall be based on the mean deceleration calculated over the stopping distance. The stopping distance shall be the distance covered by the tractor from the moment when the driver begins to actuate the control of the device until the moment when the tractor stops.

E: The stopping distance shall be the distance covered by the tractor from the moment when the driver begins to actuate the control of the device until the moment when the tractor stops.

e) Opinion questions: question asking for the opinions/feelings/ideas about people, topics, events...

Examples:

Q: *Which is the European Parliament position with respect to terrorism?*

P: Today, we in the European Parliament would like to speak out loudly and clearly against the indiscriminate violence of terrorism. We condemn utterly the senseless destruction of human life, the deaths of entire families as a result of blind fanaticism, which causes people to kill their fellow human beings and to trample human dignity underfoot. Terrorism is a direct attack on freedom, human rights and democracy. Terrorism is an attempt to destroy by means of indiscriminate violence the values which unite us in the European Union and within our Member States.

E: loudly and clearly against the indiscriminate violence of terrorism

Q: *Which is the opinion of the Committee on Employment and Social Affairs about social cohesion?*

P: Mr President, Mr Barroso, it is the unanimous opinion of the Committee on Employment and Social Affairs that I should like to share with you this morning, for we are looking to see a real promotion of social cohesion in this recovery plan. Social cohesion means being integrated into the labour market. To begin with, then, we want to keep all employees in their jobs and get the unemployed back to work by, among other things, directing the Globalisation Adjustment Fund towards new training courses, so that the workforce is prepared for when we emerge from the crisis.

E: we are looking to see a real promotion of social cohesion

Note 1: Each exact answer is **one contiguous substring** taken from the supporting paragraph.

Note 2 Exact responses are shown as an illustration only. In the case of more experimental question types like procedure or opinion it may not be clear what the exact answer is. The purpose of demarcating it is to show to the evaluators what the question creators considered to be the exact answer. In all cases the exact answer returned by a system will be judged on its own merits.

f) Other questions: The final category is for any reasonable questions which you would like to set and which do not fall into the other categories. Feel free to experiment with new ideas.

Since we aim at investigating systems' performance in answering all the above mentioned categories of questions, the test sets must be balanced, which means they should cover all the categories instead of addressing just a few of them. Thus the distribution of the questions should be 15/75 factoids and 12/75 for each of the other five categories.

Moreover, besides the ACQUIS collection, also documents from Europarl must be used to create questions. The proportion of questions should indicative be the following:

Acquis: 10/75 factoids and 8/75 for each of the other five categories;

Eurparl: 5/75 factoids and 4/75 for each of the other five categories.

N.B. Temporal restriction: all types of questions may contain a temporal restriction, but this is not a mandatory requirement.

IV. RESPONSES

To support both the Paragraph Selection Task and the Answer Selection Task, we need to know for each question, the paragraph containing the answer and the substring of it which comprises (in the view of the question generator) the nearest “exact” answer.

As noted above, exact answers for traditional factoid questions are well understood and unambiguous, while exact answers for other question types may be open to some debate. Nevertheless, it will be extremely useful to assessors to be able to see a reference exact answer to use as a starting point in assessing a system response.

All questions must be answered within **one single paragraph** containing the relevant information. **Paragraphs are delimited by the mark <p>** in the collections. Each has a separate ID.

- The paragraph must not only contain the correct answer, but must also contain the complete context required to support that answer as being correct.

Each paragraph returned must be associated to:

- the id of the paragraph [*p_id*] from which the passage has been cut/pasted in the source language ACQUIS / EuroParl collection.
- the id of the document [*doc_id*] of the ACQUIS / EuroParl aligned corpus from which the paragraph has been extracted;

For the sake of completeness also the corresponding English paragraph – together with its *p_id* – must be cut and pasted.

N.B.: Please make sure that also in the English collection the answer is contained in one single paragraph.

The exact answer, as discussed above, is exactly one substring of the chosen paragraph. The shortest meaningful substring is likely to be one word while the longest possible substring will be the entire paragraph.

In certain cases of e.g. procedure or opinion questions it is quite possible that the exact answer will in fact be the entire paragraph.

V. SUGGESTIONS BASED ON LAST YEAR

Last year was the first one working with Acquis. Based on an analysis of the questions created last year, we have identified two major problems:

Firstly, questions which were **too vague** or were **unclear**. Of course, the corpus is not precise or clearly worded and this makes our task very difficult. Here are examples from last year:

Q18. Which kinds of materials can be sold in a country that is part of the Union?

Whereas, therefore, only propagation material of varieties whose cultivation is authorised in a Member State should be marketed without restriction in that State;

‘materials’ can mean many different things. In a dialogue we would probably respond “what kind of materials” whereupon you could respond “propagation materials used in horticulture”. The key to solving this problem therefore is to specify the materials more precisely in the question and to ensure that the supporting paragraph contains that specification. In the above example it does not, so we would have to reject the paragraph and think of another question.

Q33. Which techniques used for obtaining certain products do not require a special marking?

- there are no safety reasons for labelling which states that the product has been obtained by genetic modification techniques,

Once again, words like ‘techniques’ and ‘products’ are extremely general and cannot pin down a particular answer in a large corpus.

Q53. What is one of the objects of control in member states in favour of the customers?

Whereas one of the objects of controls in each Member State is to give assurance to customers that quantities delivered correspond to the price paid ; whereas consequently the aim of this Directive is not to abolish these controls but to eliminate differences between rules in so far as these constitute a hindrance to trade;

Here, ‘control’ is a very general term. No doubt the concept and scope of controls has been discussed earlier in the document, but unfortunately we are not allowed to rely on this.

Q226. What kind of cover do oil tankers need?

(15) Accidental pollution by oil is likely to cause considerable damage to the environment and the economy of the region concerned. It is therefore necessary to verify whether oil tankers calling at European Community ports have appropriate cover for such risks.

The word 'cover' can mean many things. On a ship the most likely meaning could be the steel covers which go over each hold! We could have said something like 'What kind of insurance cover to oil tankers need when they will call into EU ports?'

The second problem encountered last year was that questions were not supported by the reference document. In circuitously worded texts such as we are working with, it is necessary to be very careful about this point. Here are some examples:

Q2. What does the creation of the council directive on the education of the migrating workers' children aim at?

Whereas in its resolution of 21 January 1974 concerning a social action programme (3), the Council included in its priority actions those designed to improve the conditions of freedom of movement for workers relating in particular to reception and to the education of their children;

The paragraph does not really answer question as the council directive on education is not mentioned.

Q12. What are the Community's procedures for contributing to the struggle against the producers of narcotics?

Whereas, in the spirit of the United Nations Convention, the Community should contribute to efforts to combat drug trafficking by producer countries ; whereas, in this connection, special machinery should be introduced for (1) OJ No L 144, 2.6.1981, p. 1. (2) OJ No L 90, 2.4.1987, p. 3. monitoring the products listed in Table II of the Annex where they are the subject of trade with those countries, despite the fact that in general these products give rise to much lawful trade ; whereas the cooperation of the countries concerned should be sought to ensure better monitoring of such trade;

The paragraph does not mention any procedures. Probably these are later in the document?

Q60. Why are technical requirements for unalloyed steel gas cylinders with capacity up to 150 litres necessary?

Whereas Council Directive 76/767/EEC of 27 July 1976 on the approximation of the laws of the Member States relating to common provisions for pressure vessels and methods of inspecting them (4), lays down in particular the procedures of EEC pattern approval and verification for these vessels ; whereas according to this Directive it is advisable to lay down the technical requirements to be complied with by EEC-type welded unalloyed steel gas cylinders with a capacity of 0,5 to 150 litres in order to be put into circulation, commercialized and used without restraint after having undergone inspection and bearing accordingly a mark and a symbol,

The paragraph does not clearly answer the question. It says the requirements are "advisable" not "necessary", and the reason is just to put the cylinders into use, which is a tautology.

Q197. When can tolls be set on motorways?

2. (a) Tolls and user charges shall be imposed only on users of motorways or other multi-lane roads with characteristics similar to motorways, or users of bridges, tunnels and mountain passes.

The text does not answer the question. It just says they can be imposed.

Q322. Why should national regulatory authorities gather information from market players?

(13) National regulatory authorities need to gather information from market players in order to carry out their tasks effectively. Such information may also need to be gathered on behalf of the Commission, to allow it to fulfil its obligations under Community law. Requests for information should be proportionate and not impose an undue burden on undertakings. Information gathered by national regulatory authorities should be publicly available, except in so far as it is confidential in accordance with national rules on public access to information and subject to Community and national law on business confidentiality.

This 'why' question is tautological because the answer is to carry out tasks effectively. That is surely the reason for doing anything at all.

Here are some examples of good questions from last year:

Q434. Which is the Brockmann procedure to standardise aluminium oxide?

3.9. Aluminium oxide, standardized according to Brockmann : ash for 8 hours at 750 Å°C, cool in a desiccator and keep in a brown glass bottle fitted with a ground-glass stopper. Before use in chromatography moisten as follows : place in a brown glass bottle 10 g aluminium oxide and 0 77 ml water, seal with a stopper, reheat for 5 minutes in a boiling water bath while shaking. Leave to cool. Verify the activity of the aluminium thus prepared by subjecting a known quantity of retinol (3.17) (ca. 500 IU) to the procedure of 5.3 and 5.4 and checking recovery.

This is a nice procedure as there are some clear steps involved.

Q473. What is the purpose of the vehicle identification number?

THE VEHICLE IDENTIFICATION NUMBER IS A FIXED COMBINATION OF CHARACTERS ASSIGNED TO EACH VEHICLE BY THE MANUFACTURER . ITS PURPOSE IS TO ENSURE THAT EVERY VEHICLE CAN BE CLEARLY IDENTIFIED OVER A PERIOD OF 30 YEARS THROUGH THE INTERMEDIARY OF THE MANUFACTURER , WITHOUT THERE BEING A NEED FOR REFERENCE TO FURTHER DATA . THE VEHICLE IDENTIFICATION NUMBER MUST CONFORM TO THE FOLLOWING REQUIREMENTS :

This is a good example of a clear purpose question with a clear answer in the supporting paragraph.

One final point. When a procedure is asked for and when a reference to a Regulation is given as answer this is considered unsupported. The procedure must be explicitly listed in the supporting paragraph.