# UNED at CL-SR CLEF 2005: Mixing Different Strategies to Retrieve Automatic Speech Transcriptions

Fernando López-Ostenero, Víctor Peinado and Valentín Sama
NLP Group, ETSI Informática, UNED
c/ Juan del Rosal, 16, E-28040 Madrid, Spain
{flopez, victor, vsama}@lsi.uned.es

**Abstract**

In this paper we describe the UNED's participation in the CLEF CL-SR 2005 track. We have tested several strategies to clean the automatic transcriptions and we have performed 84 different runs mixing these strategies with a proper noun recognition and different pseudo-relevant feedback approaches, in order to study the influence of each method in the retrieval process both in monolingual and cross-lingual environments.

We noticed that the influence of proper noun recognition is higher on the cross-lingual environment, where MAP scores double when we use our entity recognizer. The best pseudo-relevance feedback approach was the one using the MANUALKEYWORDS field. The effects of the different cleaning strategies were very similar, except for character trigrams, which obtained poor scores compared with the full word approaches.

## Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval

## General Terms

speech recognition, pseudo-relevance feedback

## Keywords

speech recognition, named entities recognition, cross-language information retrieval

# 1 Introduction

The goal of the CLEF CL-SR 2005 track is to develop and evaluate systems for ranked retrieval of spontaneous conversational speech, over a collection of 8104 segments of interviews with different people.

Our participation in the track is focused on testing and mixing different techniques to improve the retrieval effectiveness: strategies to clean the documents, proper-noun recognition and different pseudo-relevance feedback approaches.

The effects of the cleaning strategies are very similar for all methods using full words (*morpho*, *pos* and *clean*). The manual keywords show to be the best field to use in a pseudo-relevance feedback approach. When using our entity recognizer, we also improve the MAP scores in both monolingual and cross-lingual environments. However, we also noticed that the influence of proper noun detection is bigger on the cross-lingual environment, because we use the entity recognizer to detect nouns that should not be translated.

The remaining sections of this paper are divided as follows: in Section 2 we describe out testbed, the design of our submitted runs, and the new strategies used in our additional experiments. In Section 3 we present the results of our 84 runs and we analyze the influence of: proper noun recognition in monolingual and cross-lingual environments (3.1), cleaning strategies (3.2) and the different pseudo-relevance feedback methods used (3.3). Finally, in Section 4 we present some conclusions.

# 2 Experiment design

Following the CL-SR CLEF 2005 guidelines we submitted five different runs, and we perform several other experiments after receive the results and the relevance assessments.

## 2.1 Testbed

The test collection consists of 8104 segments from interviews of Holocaust survivors. Each document has several fields with different pieces of information about the segment. In our experiments we have used the following information:

- We compared the two different automatic transcriptions (ASR2003A and ASR2004A) of the segment, and we find that there are no big differences between both transcriptions. So, we decide to use only the transcription field ASR2004A.
- The human written summary of the segment (SUMMARY) to detect Proper Nouns which not need to be translated.
- Three different sets of keywords: one set of manually selected keywords (MANUALKEYWORD) and two different automatic sets of keywords (AUTOKEYWORD2004A1 and AUTOKEYWORD2004A2).

We have tried different strategies to clean the automatic transcriptions of the documents:

1. We noticed that several words that are not recognized by the automatic transcription, in particular proper nouns, appear split in their characters, like "*l i e b b a c h a r d*". We have searched all the occurrences of a list of single characters into the documents, and we have joined them assuming that these characters form a whole unrecognized word.
2. When one speaks, it is usual to repeat some words to emphasize a specific part of the conversation or to show to the other person a thinking process ("*let me think, yes, yes, yes...*"). Automatic Speech Recognizers transcribe these duplicated words and when performing a retrieval process, the results can be affected by these words. So we have removed all extra occurrences of the duplicated words.

   The resulting documents after these two first steps have been indexed in a collection that we will refer as **clean**.
3. In Information Retrieval, the most informative words seems to be nouns, adjectives and verbs. Our next step was to clean the documents, removing all words except nouns, adjectives and verbs.

   We used two different approaches to perform this cleaning:

- We used the FreeLing [6] set of linguistic tools to perform a morphological analysis of the documents from the **clean** collection. The output of this analysis indicates the possible Part of Speech for a given word. If, according to this analysis, a word can act as a noun, an adjective or a verb, the word remains in the document and was removed otherwise. This technique was used successfully to extract noun-phrases from a Spanish collection [3] and it proved to be very useful in an Information Retrieval environment. We built a new collection called **morpho** with the documents obtained after this process.
- We perform a full Part of Speech tagging using freeling again. This process is more complex than the previous one, and includes a PoS Disambiguation phase in order to select only one of the possible Part of Speech for a given word. Like in the previous process, only words that act as nouns, adjectives or verbs remain in the resulting documents, which have been indexed in a collection that we will refer as **pos**.

4. Additionally, we split the cleaned documents in character 3-grams to compare the performance of this simple approach with the performance of more complex cleaning process. This collection will be referred as **3grams**.

We have used the English and the Spanish topics, provided by the organizers, from those we have only used TITLE and DESCRIPTION fields in all our runs. For each topic we have removed the usual stopwords and, in the 3-grams runs, we have split each word into character 3-grams individually.

For our cross-lingual runs, we used a query translation approach following Pirkola's proposal [4], where alternative translations for a term were taken as synonyms, giving them equal weights.

Finally we have used INQUERY [2] as a search engine.

### 2.2 Submitted runs

Following the CL-SR CLEF guidelines, we submit five different runs:

1. A monolingual run using the **3grams** collection and the English topics expressed as 3-grams (mono-3grams).
2. A monolingual run using the **morpho** collection (mono-morpho).
3. A cross-lingual run using the **morpho** collection and the Spanish topics translated into English (trans-morpho).
4. A monolingual run using the **pos** collection (mono-pos).
5. A cross-lingual run using the **pos** collection and the Spanish topics translated into English (trans-pos).

The results of the submitted runs are shown in Table 1, where we can compare our results with the best monolingual and Spanish cross-lingual runs.

Regarding these numbers we can draw some preliminary conclusions:

- Our runs are far from the best monolingual and Spanish cross-lingual runs, so there is room for improvement.
- MAP scores of **morpho** and **pos** runs are very similar in both monolingual and cross-lingual environments. As expected **pos** scores slightly better than **morpho**, but with only two different runs we don't have enough data to conclude that the PoS Disambiguation helps to clean these documents.
- Character 3-grams run scores worse than full word retrieval (75.6%). Again we have too few data to conclude that it's better to use full word retrieval than a 3-grams approach.

| Ranking | MAP | Run | Language |
|---|---|---|---|
| 1 | 0.3129 | UMD (best English run) | English (monolingual) |
| 5 | 0.1863 | University of Ottowa (best Spanish run) | Spanish (cross-lingual) |
| 20 | 0.0934 | mono-pos | English (monolingual) |
| 21 | 0.0918 | mono-morpho | English (monolingual) |
| 29 | 0.0706 | mono-3grams | English (monolingual) |
| 32 | 0.0373 | trans-pos | Spanish (cross-lingual) |
| 33 | 0.0370 | trans-morpho | Spanish (cross-lingual) |

Table 1: Comparison of results of submitted runs

- Just using a bilingual dictionary and a Pirkola's approach, our cross-lingual runs reach 40% MAP of their respective monolingual counterparts. In some cases (see section 3.1) a bad translation of some proper nouns difficult the cross-lingual search.

With only five different runs it's very difficult to obtain clear conclusions. According the suggestion of CL-SR CLEF organizers, we have run more experiments after the official submission. These experiments are described in the next section.

### 2.3 Additional experiments

With these additional experiments, our intention was to test the effects of two new strategies (proper noun identification and pseudo-relevance feedback) and compare all possible combination of all our different approaches.

#### Proper noun identification

We have used our entity recognizer [5] in order to improve the query structure identifying possible proper nouns in the topics. These entities have been used in different ways in our monolingual and cross-lingual experiments:

- **Monolingual**: we just identify the proper nouns contained in the topics and, using the *#phrase* operator of INQUERY, we have structured the query.
- **Cross-lingual**: if we use the same strategy as above, the identified proper nouns in Spanish topics maybe will not appear in the English documents. For instance, in the topic **#1133** ("*The story of Varian Fry*"), we identify these proper nouns:
  - Varian Fry
  - Comité de Rescates de Emergencia
  - Marsella

  We used the recognizer in order to identify possible proper nouns in the SUMMARY field of the documents too. Only proper nouns that appear in both lists have been used to structure the query.

  In the given example, the only proper noun that appear in the SUMMARY field of the documents was "*Varian Fry*". So our strategy was to translate the topic, except these names.

Once we have the proper nouns identified, we use the INQUERY's operator *#phrase* with each one to structure the query. Below we can see the topic #2012 unstructured and structured using the proper noun identified on it.

- **Topic #2012 unstructured**: #sum( collaboration local population information collaboration local population german authorities east central europe holocaust );

- **Topic #2012 structured**: #sum( #phrase(German Authorities) #phrase(East Central Europe) #phrase(Holocaust) collaboration local population information collaboration local population );

### Relevance Feedback

We decide to test a Pseudo-Relevance Feedback [1] (prf) approach to check the utility of the keyword fields of the documents. We built five different relevance feedback methods:

- A collection (AK1) using the AUTOKEYWORD2004A1 field.
- A collection (AK2) using the AUTOKEYWORD2004A2 field.
- A collection (AK12) mixing the keywords present in both autokeyword fields.
- A collection (MK) using the MANUALKEYWORD field.
- A collection (MKAK12) mixing the keywords from the three keyword fields.

In order to mix the keywords from different fields, we used the following process:

1. Each automatic keyword was scored according to his order of appearance in the field: the first keyword obtains a score of 20 (there are 20 keywords in each autokeyword field), and the last one just 1.
2. If a keyword appears in both autokeyword fields, his final score was the sum of the scores obtained in each field.
3. In order to create the AK12 collection, we selected the 20 keywords with higher score.
4. When building the MKAK12 collection, we first selected the $n$ manual keywords, and then, we added the $20 - n$ first keywords from AK12 collection.

When searching, we select the top 10 ranked documents and we combine their keywords using the same method as described above to mix the keywords from different fields. Finally we refine the query adding the top 20 keywords.

### Full set of runs

Our intention was to test all possible combinations of each feature: topic language, proper noun identification, cleaning method and relevance feedback. However, we notice that the meaning of cross-lingual runs using 3-grams might be confusing. The translated query is structured with the *#syn* operator of INQUERY. If we split the translated query into 3-grams, we can't use the *#syn* operator to structure it, because all 3-grams would be considered as synonyms even if they came from the same word. So we have not used 3-grams on our cross-lingual runs.

$$\underbrace{\begin{pmatrix}\text{mono}\\ \text{trans}\end{pmatrix}}_{\textit{language}} \times \underbrace{\begin{pmatrix}\text{noent}\\ \text{ent}\end{pmatrix}}_{\textit{proper noun}} \times \underbrace{\begin{pmatrix}\text{3grams}\\ \text{clean}\\ \text{morpho}\\ \text{pos}\end{pmatrix}}_{\textit{cleaning method}} \times \underbrace{\begin{pmatrix}\text{NO}\\ \text{AK1}\\ \text{AK2}\\ \text{AK12}\\ \text{MK}\\ \text{MKAK12}\end{pmatrix}}_{\textit{relevance feedback}}$$

Figure 1: Combination of all features

Each run is named with the labels of the different features used on it. For instance "*mono-noent-morpho-AK2*" represents a monolingual run without proper noun identification, over the

**morpho** collection and using the AUTOKEYWORD2004A2 field for the Pseudo-Relevance Feedback process.

On figure 1 we can see all possible values for each feature: 96 different combinations. But, if we exclude the cross-lingual runs when using 3-grams, there are 84 different runs.

# 3 Results and discussions

The results of all our runs are shown in Table 2.

We can draw some preliminary points:

- Our best run, *mono-ent-morpho-MK*, scores a 25.95% MAP, a 82.9% of the best submitted monolingual run, from the University of Maryland. We have obtained a 277.8% of improvement respect our best submitted monolingual run.
- Our best cross-lingual run *trans-ent-pos-MK* scores about a 131.2% MAP respect the best submitted Spanish run from the University of Ottowa. In this case, the improvement respect our best submitted cross-lingual run is 545.8%.
- The best strategy seems to be pseudo-relevance feedback using the MANUALKEYWORDS field, followed by the combination of the three keywords fields.
- The monolingual 3-grams runs score poorly, reaching only a 30% MAP of our best run.

Let analyze more carefully the influence of the different approaches:

## 3.1 Language and Proper noun effects

On table 3 we can compare the effects of the proper noun detection in both monolingual and cross-lingual runs. The numbers on the *ent* and *noent* columns show the percentage of the MAP of the cross-lingual runs compared with the MAP of the monolingual. The numbers on the *mono* and *trans* represent the increment of the MAP when using proper noun detection technique.

Regarding these numbers we can infer some interesting points:

- Using proper nouns, the MAP of cross-lingual runs reach 75% of the monolingual runs. Without proper nouns, cross-lingual runs reach only $35-40\%$ MAP of monolingual runs.
- Influence of proper nouns is higher on cross-lingual runs, increasing MAP more than twice respect to *noent* runs. On monolingual runs the increment is worthless and, probably, statistically not relevant.
- The influence of proper nouns is also worthless on 3-grams runs, even the best 3-grams run is *mono-noent-3grams-MK* that not uses proper noun detection.

For instance, on topic *#1113* ("*The story of Varian Fry*"), the influence of proper noun detection is very important, because in Spanish the word "Varian" can be identified as a verbal form of "Variar" (to vary, change) and is wrongly translated:

- **trans-noent-clean-NO**: #sum( #syn( depart motley variegate vary deviate diverge fluctuate alter change ) fry #syn( depart motley variegate vary deviate diverge fluctuate alter change ) fry #syn( commission board committee ) #syn( ransom rescue deliver redeem recover ) #syn( ransom salvage rescue ) #syn( yard mil g thou k grand m thousandth chiliad thousand 1000 ) #syn( living life sprightliness livelihood lifespan spirit liveliness cheer lifetime cheerfulness ) );
- **trans-ent-clean-NO**: #sum( #phrase(Varian Fry) #phrase(Varian Fry) historia historiar #syn( commission board committee ) rescatar rescate #syn( egression exigency emergency pinch egress growth emergence urgency ) salvar salvo salvar #syn( yard mil g thou k grand m thousandth chiliad thousand 1000 ) #syn( living life sprightliness livelihood lifespan spirit liveliness cheer lifetime cheerfulness ) #syn( marseille marseilles ) );

| MAP | R-PREC | Experiment |
|---|---|---|
| 0.2595 | 0.3046 | mono-ent-morpho-MK |
| 0.2583 | 0.3001 | mono-ent-pos-MK |
| 0.2557 | 0.3025 | mono-ent-clean-MK |
| 0.2499 | 0.2879 | mono-noent-morpho-MK |
| 0.2498 | 0.2873 | mono-noent-pos-MK |
| 0.2462 | 0.2860 | mono-noent-clean-MK |
| 0.2396 | 0.2897 | mono-ent-pos-MKAK12 |
| 0.2353 | 0.2855 | mono-ent-morpho-MKAK12 |
| 0.2299 | 0.2895 | mono-ent-clean-MKAK12 |
| 0.2284 | 0.2740 | mono-noent-pos-MKAK12 |
| 0.2245 | 0.2711 | mono-noent-morpho-MKAK12 |
| 0.2224 | 0.2774 | mono-noent-clean-MKAK12 |
| 0.2036 | 0.2444 | trans-ent-pos-MK |
| 0.2000 | 0.2475 | trans-ent-clean-MK |
| 0.1982 | 0.2437 | trans-ent-morpho-MK |
| 0.1931 | 0.2443 | trans-ent-pos-MKAK12 |
| 0.1880 | 0.2420 | trans-ent-morpho-MKAK12 |
| 0.1853 | 0.2411 | trans-ent-clean-MKAK12 |
| 0.1025 | 0.1465 | trans-noent-morpho-MK |
| 0.1016 | 0.1421 | trans-noent-pos-MK |
| 0.0994 | 0.1574 | mono-noent-pos-AK12 |
| 0.0991 | 0.1597 | mono-ent-pos-AK12 |
| 0.0976 | 0.1378 | trans-noent-clean-MK |
| 0.0971 | 0.1523 | mono-noent-morpho-AK12 |
| 0.0969 | 0.1562 | mono-ent-morpho-AK12 |
| 0.0953 | 0.1530 | mono-noent-clean-AK12 |
| 0.0950 | 0.1582 | mono-ent-pos-NO |
| 0.0944 | 0.1593 | mono-ent-clean-NO |
| 0.0937 | 0.1540 | mono-ent-clean-AK12 |
| 0.0935 | 0.1603 | mono-ent-morpho-NO |
| **0.0934** | **0.1522** | **mono-noent-pos-NO** |
| 0.0927 | 0.1528 | mono-noent-clean-NO |
| **0.0918** | **0.1532** | **mono-noent-morpho-NO** |
| 0.0879 | 0.1254 | trans-noent-morpho-MKAK12 |
| 0.0874 | 0.1450 | mono-ent-pos-AK2 |
| 0.0871 | 0.1431 | mono-noent-pos-AK2 |
| 0.0868 | 0.1431 | mono-ent-morpho-AK2 |
| 0.0866 | 0.1522 | mono-ent-pos-AK1 |
| 0.0865 | 0.1221 | trans-noent-pos-MKAK12 |
| 0.0860 | 0.1500 | mono-ent-morpho-AK1 |
| 0.0860 | 0.1473 | mono-noent-pos-AK1 |
| 0.0857 | 0.1225 | trans-noent-clean-MKAK12 |

| MAP | R-PREC | Experiment |
|---|---|---|
| 0.0853 | 0.1372 | mono-noent-morpho-AK2 |
| 0.0852 | 0.1468 | mono-ent-clean-AK1 |
| 0.0846 | 0.1469 | mono-noent-morpho-AK1 |
| 0.0841 | 0.1408 | mono-noent-clean-AK1 |
| 0.0837 | 0.1287 | trans-ent-pos-AK12 |
| 0.0828 | 0.1382 | mono-noent-clean-AK2 |
| 0.0827 | 0.1282 | trans-ent-morpho-AK12 |
| 0.0826 | 0.1423 | mono-ent-clean-AK2 |
| 0.0789 | 0.1282 | trans-ent-clean-AK12 |
| 0.0780 | 0.1134 | mono-noent-3grams-MK |
| 0.0769 | 0.1370 | trans-ent-pos-AK1 |
| 0.0766 | 0.1361 | trans-ent-morpho-AK1 |
| 0.0752 | 0.1329 | trans-ent-clean-AK1 |
| 0.0740 | 0.1127 | mono-ent-3grams-MK |
| 0.0735 | 0.1202 | trans-ent-morpho-NO |
| 0.0731 | 0.1193 | trans-ent-pos-NO |
| 0.0731 | 0.1175 | trans-ent-clean-NO |
| 0.0725 | 0.1198 | trans-ent-pos-AK2 |
| 0.0717 | 0.1191 | trans-ent-morpho-AK2 |
| 0.0715 | 0.1196 | trans-ent-clean-AK2 |
| **0.0706** | **0.1119** | **mono-noent-3grams-NO** |
| 0.0650 | 0.1029 | mono-ent-3grams-MKAK12 |
| 0.0649 | 0.1125 | mono-noent-3grams-MKAK12 |
| 0.0601 | 0.1020 | mono-ent-3grams-NO |
| 0.0541 | 0.0892 | mono-ent-3grams-AK12 |
| 0.0475 | 0.0870 | mono-ent-3grams-AK1 |
| 0.0427 | 0.0757 | mono-ent-3grams-AK2 |
| 0.0423 | 0.0850 | mono-noent-3grams-AK12 |
| 0.0411 | 0.0838 | mono-noent-3grams-AK1 |
| 0.0393 | 0.0667 | mono-noent-3grams-AK2 |
| **0.0373** | **0.0750** | **trans-noent-pos-NO** |
| 0.0372 | 0.0746 | trans-noent-clean-NO |
| **0.0370** | **0.0759** | **trans-noent-morpho-NO** |
| 0.0346 | 0.0724 | trans-noent-pos-AK1 |
| 0.0346 | 0.0687 | trans-noent-morpho-AK1 |
| 0.0343 | 0.0713 | trans-noent-pos-AK12 |
| 0.0342 | 0.0723 | trans-noent-clean-AK1 |
| 0.0331 | 0.0673 | trans-noent-morpho-AK12 |
| 0.0326 | 0.0634 | trans-noent-clean-AK12 |
| 0.0290 | 0.0664 | trans-noent-morpho-AK2 |
| 0.0288 | 0.0663 | trans-noent-pos-AK2 |
| 0.0282 | 0.0673 | trans-noent-clean-AK2 |

Table 2: Results of all runs (submitted runs in boldface)

| | trans/mono | | ent/noent | |
|---|---|---|---|---|
| Experiment | ent | noent | mono | trans |
| clean-NO | 77.4% | 40.1% | 101.83% | 196.50% |
| clean-AK1 | 88.3% | 40.7% | 101.30% | 219.88% |
| clean-AK2 | 86.6% | 34.1% | 99.75% | 253.54% |
| clean-AK12 | 84.2% | 34.2% | 98.32% | 242.02% |
| clean-MK | 78.2% | 39.6% | 103.85% | 204.91% |
| clean-MKAK12 | 80.6% | 38.5% | 103.37% | 216.21% |
| pos-NO | 76.9% | 39.9% | 101.71% | 195.97% |
| pos-AK1 | 88.8% | 40.2% | 100.69% | 222.25% |
| pos-AK2 | 83.0% | 33.1% | 100.34% | 251.73% |
| pos-AK12 | 84.5% | 34.5% | 99.69% | 244.02% |
| pos-MK | 78.8% | 40.7% | 103.40% | 200.39% |
| pos-MKAK12 | 80.6% | 37.9% | 104.90% | 223.23% |
| morpho-NO | 78.6% | 40.3% | 101.85% | 198.64% |
| morpho-AK1 | 89.1% | 40.9% | 101.65% | 221.38% |
| morpho-AK2 | 82.6% | 34.0% | 101.75% | 247.24% |
| morpho-AK12 | 85.3% | 34.1% | 99.79% | 249.84% |
| morpho-MK | 76.4% | 41.0% | 103.84% | 193.36% |
| morpho-MKAK12 | 79.9% | 39.2% | 104.81% | 213.87% |

Table 3: Influence of proper noun detection, comparison of monolingual and cross-lingual runs

### 3.2 Cleaning effects

Regarding the influence of the cleaning method we can conclude that the best cleaning strategy seems to be *morpho*, but the differences respect *pos* and *clean* are minimal and, probably statistically not relevant.

Using just a morphological analyzer to identify possible part of speech for a given word, proves to be a very useful strategy in information retrieval if we don't have a full part of speech tagger.

Again, character 3-grams show to be a bad cleaning strategy when compared with full words approaches.

### 3.3 Relevance feedback

On table 4 we can compare the differences between the different pseudo-relevant feedback strategies tested. Each column represent the MAP percentage of one prf method respect to another. For instance MKAK12/MK column represent the percentage of MKAK12 prf respect MK prf.

| Experiment | MKAK12/MK | AK12/MKAK12 | AK1/AK2 | MK/NO |
|---|---|---|---|---|
| mono-ent-clean | 89.91% | 40.75% | 103.14% | 270.86% |
| mono-ent-pos | 92.76% | 41.36% | 99.08% | 271.89% |
| mono-ent-morpho | 90.67% | 41.18% | 99.07% | 277.54% |
| mono-noent-clean | 90.33% | 42.85% | 101.57% | 265.58% |
| mono-noent-pos | 91.43% | 43.52% | 98.73% | 267.45% |
| mono-noent-morpho | 89.83% | 43.25% | 99.17% | 272.22% |
| trans-ent-clean | 92.65% | 42.57% | 105.17% | 273.59% |
| trans-ent-pos | 94.84% | 43.34% | 106.06% | 278.52% |
| trans-ent-morpho | 94.85% | 43.98% | 106.83% | 269.65% |
| trans-noent-clean | 87.80% | 38.03% | 121.27% | 262.36% |
| trans-noent-pos | 85.13% | 39.65% | 120.13% | 272.38% |
| trans-noent-morpho | 85.75% | 37.65% | 119.31% | 277.02% |

Table 4: Influence of relevance feedback methods

- The best prf method is MK (average increment of MAP using the prf over manual keywords field respect no relevance feedback is about 271.6%), nearly followed by MKAK12 (an average MAP of 90.5% respect MK).
- There are no big differences between the use of each automatic keyword field, but prf using AUTOKEYWORD2004A1 field seems to obtain high MAP score. And, when combining both fields (AK12), MAP scores 41.51% of MKAK12 on average.

# 4 Conclusions and future work

In this paper, we have shown different techniques to improve retrieval of automatic speech transcriptions in both monolingual and cross-lingual environments.

- We have tested four different cleaning techniques. Differences between full word techniques (*clean*, *morpho* and *pos*) are worthless, but a character 3-grams approach seems to be worse.
- Pseudo-relevance feedback using manually generated keywords shows to be the best option to increment the performance of the retrieval, with an average percentage of 271.6% respect no relevance feedback.
- The use of an entity recognizer to identify proper nouns, proves to be very useful, specially on a cross-lingual environment, where the MAP scores twice when using them.

Our intention is to perform further analysis over the results, including statistical relevance tests to determine the influence of the different methods we have tried.

We also want to test a different approach to identify proper nouns in the automatic transcriptions, or in the automatic keyword fields, instead of using the manual summary of the documents. Maybe the big improvement detecting proper nouns in the cross-lingual environment is due to the use of a manually generated field, similarly to the best scores obtained when using MANUALKEYWORDS field in pseudo-relevance feedback.

# Acknowledgments

This work has been partially supported by the Spanish Government under project R2D2-Syembra (TIC2003-07158-C04-02). Víctor Peinado holds a PhD grant by UNED (Universidad Nacional de Educación a Distancia).

# References

[1] C. Buckley and G. Salton and J. Allan and A. Singhal. Automatic Query Expansion Using SMART: TREC 3. In *Proceedings of the 3rd Text Retrieval Conference (TREC3)*, pages 69–80. National Institute of Standards and Technology (NIST), Gaithesburg, MD, 1995.

[2] J.P. Callan and W.B. Croft and S.M. Harding. The Inquery Retrieval System. In *Proceedings of the Third International Conference on Database and Expert Systems Applications*, pages 78–83. Springer-Verlag, 1992.

[3] A. Peñas. *Website Term Browser: Un sistema interactivo y multilingüe de búsqueda textual basado en técnicas lingüísticas* . PhD thesis, Departamento de Lenguajes y Sistemas Informáticos, Universidad Nacional de Educación a Distancia, 2002.

[4] A. Pirkola. The Effects of Query Structure and Dictionary Setups in Dictionary-Based Cross-Language Information Retrieval. In *Proceedings of SIGIR'98, 21st ACM International Conference on Research and Development in Information Retrieval*, pages 55–63, 1998.

[5] V. Peinado and F. López-Ostenero and J. Gonzalo. UNED at ImageCLEF 2005: Automatically Structured Queries with Named Entities over Metadata. In *Cross Language Evaluation Forum, Working Notes for the CLEF 2005 Workshop*, 2005.

[6] X. Carreras and I. Chao and L. Padró and M. Padró. FreeLing: An Open-Source Suite of Language Analyzers. Technical report.