

Interactive Cross-Language Searching: phrases are better than terms for query formulation and refinement

Fernando López-Ostenero, Julio Gonzalo, Anselmo Peñas and Felisa Verdejo*

Abstract

This paper summarizes the participation of the UNED group in the CLEF 2002 Interactive Track. We focused on interactive query formulation and refinement, comparing two approaches: a) a reference system that assists the user to provide adequate translations for terms in the query; and b) a proposed system that assists the user to formulate the query as a set of relevant phrases, and to select promising phrases in the documents to enhance the query. All collected evidence indicates that the phrase-based approach is preferable: the official $F_{\alpha=0.8}$ measure is 65% better for the proposed system, and all users in our experiment preferred the phrase-based system as a simpler and faster way of searching.

1 Introduction

In our second participation in the CLEF Interactive track, we have focused on assisted query formulation and refinement for Cross-Language searching. Previous experiments in this area have mainly concentrated on assisted query translation [4, 1]: for every term in the query, the system displays its possible translations with some information about the meaning of each candidate (via definitions in the source language or inverse translations). The user then selects the most appropriate translations, interacting with the system to overcome translation ambiguity.

Our hypothesis to be tested in iCLEF 2002 is twofold:

- Examining translations in an (unknown) foreign language is a high-load cognitive task, and therefore it is worth exploring alternative ways of assisting cross-language query formulations.
- Selecting relevant phrases for a topic should be easier and faster than selecting translations, and phrases can be translated without human intervention more accurately than individual terms.

In order to test our hypothesis, we have used:

- An Interactive Cross-Language system that helps users to provide accurate translations for query terms, based on the information provided by a reverse dictionary. This system is used as the reference one.
- An Interactive Cross-Language system that helps users to select appropriate phrases to describe their user needs, and translates phrases in a completely automatic way.
- A common document translation strategy for both systems based on our previous iCLEF findings [3].

In Section 2, we detailed our experiment design. In Section 3, we discuss the outcome of the experiment, and in Section 4 we end up with some conclusions.

*Departamento de Lenguajes y Sistemas Informáticos, Universidad Nacional de Educación a Distancia, E.T.S.I Industriales, Ciudad Universitaria s/n, 28040 Madrid, SPAIN. flopez,julio,anselmo,felisa@lsi.uned.es

2 Experiment Design

Our experiment consists of:

- Eight native Spanish speakers with null or very low English skills.
- The Spanish version of the four official iCLEF topics.
- The English CLEF document collection (LA Times 1994).
- A reference interactive cross-language search system based on assisted term translation (System WORDS).
- A proposed system based on noun-phrase selections (System PHRASES).
- The official iCLEF latin square to combine topics, searchers and systems into 32 different searching sessions.
- The official iCLEF search procedure.

In this section we describe the most relevant aspects of the above items.

2.1 Reference system

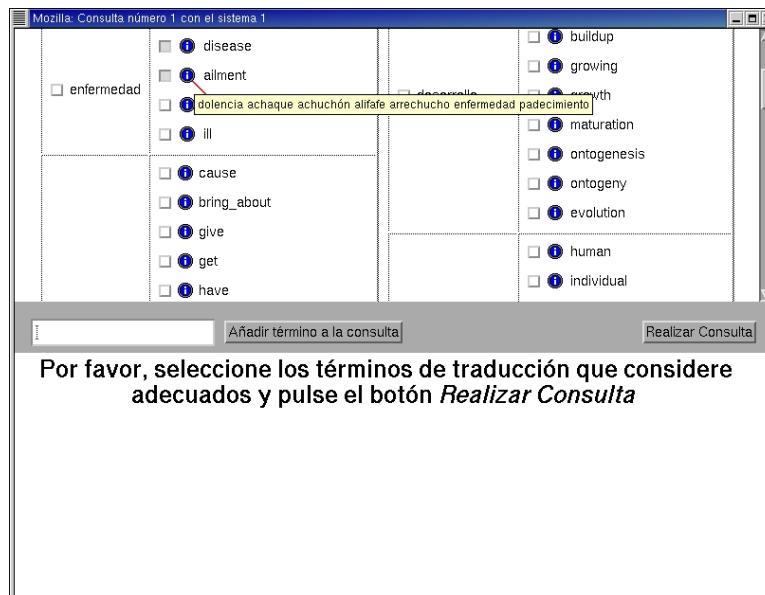
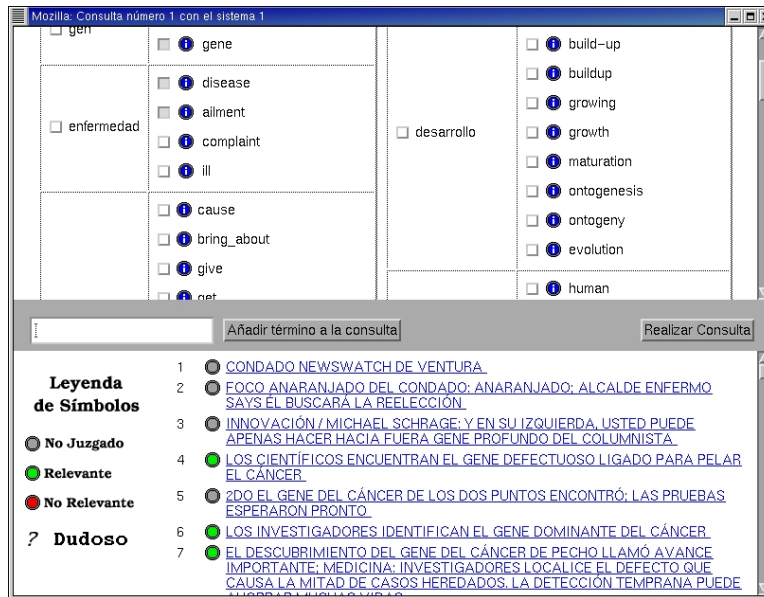


Figure 1: Reference system (WORDS), initial assisted translation

The reference system (WORDS) uses assisted query term translation and refinement all along the search process:

- **Initial query formulation.** The system translates all content words in the iCLEF topic using a bilingual dictionary, and displays possible English translations to the user. When the user points to an English term, the system displays inverse translations into Spanish. This information can be used by the searcher to decide which translations to keep and which translations to discard before performing the first search. Figure 1 illustrates this initial step.



A) Colour codes in the ranked list indicate already judged documents.



B) Clicking on a Spanish term in the document takes the user to the source English keyword matched.

Figure 2: Reference system (WORDS): visualization of rankings and single documents

- **Cross-Language search.** The system performs a monolingual search of the LA Times collection with the English terms selected by the user.
- **Ranked document list.** The ranked list of documents displays the (translated) title of the document and a colour code to indicate whether each document has already been marked as *relevant*, *not relevant* or *unsure*. Figure 2 A shows a retrieved ranked list.

- **Document selection.** Instead of using Machine Translation to display the contents of a document, the system displays a cross-language summary consisting on the translation of all noun phrases in the body of the document, plus an MT (Systran Professional 3.0) translation of the title. The user can select the document as *relevant*, mark the document as *non-relevant* or *unsure*, or leave it unmarked.
- **Query refinement by selection.** When a Spanish term in a document translation corresponds to an original English term already in the query, the user can point to the Spanish term (highlighted); then the system points to the English query term, allowing for de-selection or selection of the English term (or some of its companion translations) or the original Spanish term (then all translations are disabled). Figure 2 B illustrates this process.
- **Additional query refinement.** Additionally, the user can also enter a single term at any time along the search. Again, the system displays its possible translations into the target language, along with their inverse translations, and permits individual selection and de-selection of translations.

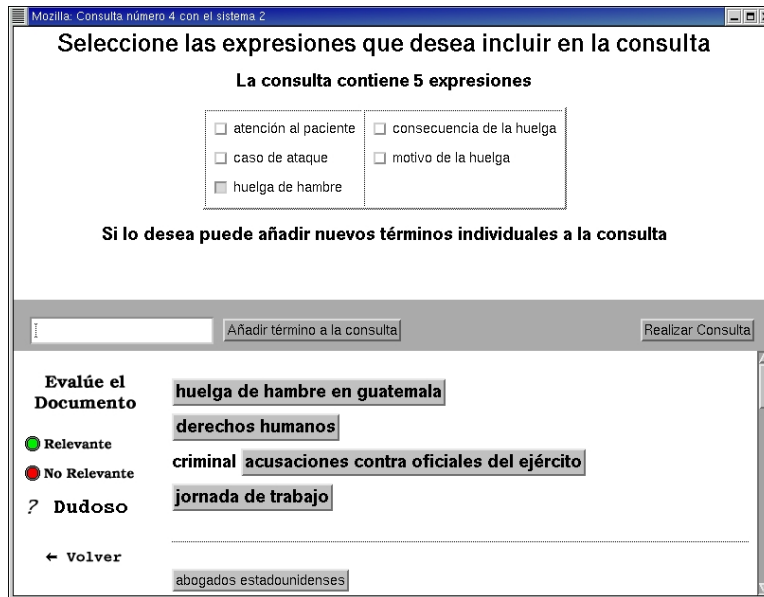
2.2 Phrase-based searching

Our proposed system uses noun phrasal information all along the Cross-Language assisted search process:

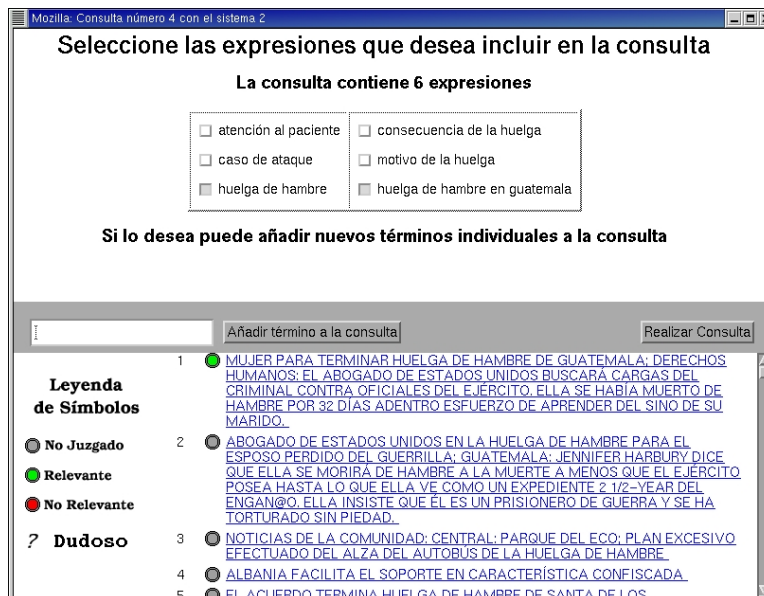
- **Initial query formulation.** The system extracts noun phrases from the full iCLEF topic, filters phrases with optimal translations, and displays the resulting set of phrases for user selection.
- **Cross-Language search.** The system translates automatically the phrases selected by the user, and performs a monolingual search in the document collection.
- **Ranked document list.** The ranked list is identical for both systems (see reference system above).
- **Document selection.** Again, document selection is identical for both systems (see WORDS system above).
- **Query refinement by term suggestion.** Optimally translated noun phrases in the documents can be selected to enrich the original query. When a user clicks on a noun-phrase in a document, the system automatically translates the noun-phrase and performs a new monolingual search with the enlarged query, updating the list of ranked documents. This process is illustrated in Figure 3.
- **Additional query refinement.** Identical in both systems (see system WORDS above).

In order to achieve such functionalities, there is a pre-processing phase using shallow Natural Language Processing techniques, which has been described in detail in [3]. The essential steps are:

- **Phrase indexing.** Shallow parsing of two comparable collections (the CLEF Spanish and English collections in this case) to obtain an index of all noun phrases in both languages and their statistics.
- **Phrase Alignment.** Spanish and English noun phrases (up to three lemmas) are aligned for translation equivalents using only a bilingual dictionary and statistical information about phrases (see [3] for details). As a result of this step, aligned phrases receive a list of candidate phrase translations in decreasing order of frequency. The result is a pseudo bilingual dictionary of phrases that is used in all other translation steps. The statistics for the CLEF English-Spanish collection can be seen in Table 1.



A) Clicking on best-aligned phrases incorporates them to the query.



B) Results of clicking the phrase “huelga de hambre en Guatemala”. The phrase is added to the query and a new ranked list is displayed.

Figure 3: Proposed system (PHRASES): query expansion by clicking a phrase in a document

- Document translation. All noun phrases are extracted and translated. Translation is performed in two steps: first, maximal aligned subphrases are translated according to the alignment information. Then, the rest of the terms are translated using an estimation that selects target terms which overlap maximally with the set of related subphrases.

Only an additional step is required at searching time:

Phrase set	Extracted	Aligned
Spanish, 2 lemmas	6,577,763	2,004,760
Spanish, 3 lemmas	7,623,168	252,795
English, 2 lemmas	3,830,663	1,456,140
English, 3 lemmas	3,058,698	198,956

Table 1: Statistics of the phrase alignment algorithm (English-Spanish CLEF collection)

- Query translation. All Spanish phrases selected by the user are replaced by: 1) the most frequent aligned English phrase and 2) the second most frequent aligned phrase, if its frequency reaches a threshold of 80% of the most frequent one. The INQUERY `phrase` operator is used to formulate the final monolingual query with all English phrases. The search is then performed using the INQUERY search engine.

2.3 Data collected

Every searcher performed 4 searches, one per iCLEF topic, alternating systems and topics according to the iCLEF latin square design. The time for each search was 20 minutes, and the overall time per searcher was around three hours, including training, questionnaires and searches (see [2] for details). For every user/topic/system combination, the following data were collected:

- The set of documents retrieved by the user, and the time at which every selection was made.
- The ranked lists produced by the system in each query refinement.
- The questionnaires filled-in by the user.
- An observational study of the search sessions.

3 Results

3.1 Official $F_{\alpha=0.8}$ scores

The official iCLEF score for both systems is $F_{\alpha=0.8}$, which combines precision and recall over the set of manually retrieved documents, favoring precision. The results of our experiment can be seen in Table 2. Our proposed system (PHRASES) improves the reference system (WORDS) by a 65% increment. In a more detailed analysis per topic, there can be seen that topic 3 was too difficult and did not contribute to the results (no searcher found relevant documents with any of the systems). All the other topics receive a better F measure with the PHRASES system than with the WORDS system. The difference is not very high for topics 1 and 2, but it is very accused for topic 4, which seemed easy for system PHRASES and very difficult for system WORDS.

The most important expression in Topic 4 is “hunger strikes” (the description is “documents will report any information relating to a hunger strike attempted in order to attract attention to a cause”). Searchers using the PHRASES system easily select “huelga de hambre” (the Spanish equivalent) from the displayed options, and the aligned translation, which is in turn “hunger strikes”, will retrieve useful documents. Searchers using the WORDS system, however, find that “huelga” (strike) and “hambre” (hunger) may receive many possible translations into English. Looking at the average F_{α} , it is obvious that they do not manage to find the appropriate translations for both terms, failing to match relevant documents.

3.2 Additional data

Besides the official F_{α} result, there are many other sources of evidence to compare both systems: additional quantitative data (time logs, ranked results for every query refinement), questionnaires

Overall $F_{\alpha=0.8}$ per system

System	$F_{\alpha=0.8}$
WORDS	.23
PHRASES	.37 (+65%)

Average $F_{\alpha=0.8}$ per topic/system

Topic	WORDS	PHRASES
1	57.25	63.5
2	28.25	30.5
3	0	0
4	0.05	55.25

$F_{\alpha=0.8}$ per topic/searcher/system combination

Searcher\Topic	1	2	3	4	Av.
1	0.62	0.44	0	0.07	0.28
2	0.55	0	0	0	0.14
3	0.62	0.35	0	0.56	0.38
4	0.64	0.49	0	0.65	0.45
5	0.66	0.15	0	0.08	0.22
6	0.62	0.14	0	0.08	0.21
7	0.56	0.51	0	0.44	0.38
8	0.56	0.27	0	0.56	0.35
Av.	0.6	0.29	0	0.31	0.3

(System WORDS in normal font, System **PHRASES** in bold font)

Table 2: Official results

filled by participants, and the observation study of their searching sessions. We discuss that additional evidence here.

3.2.1 Searching behavior across time

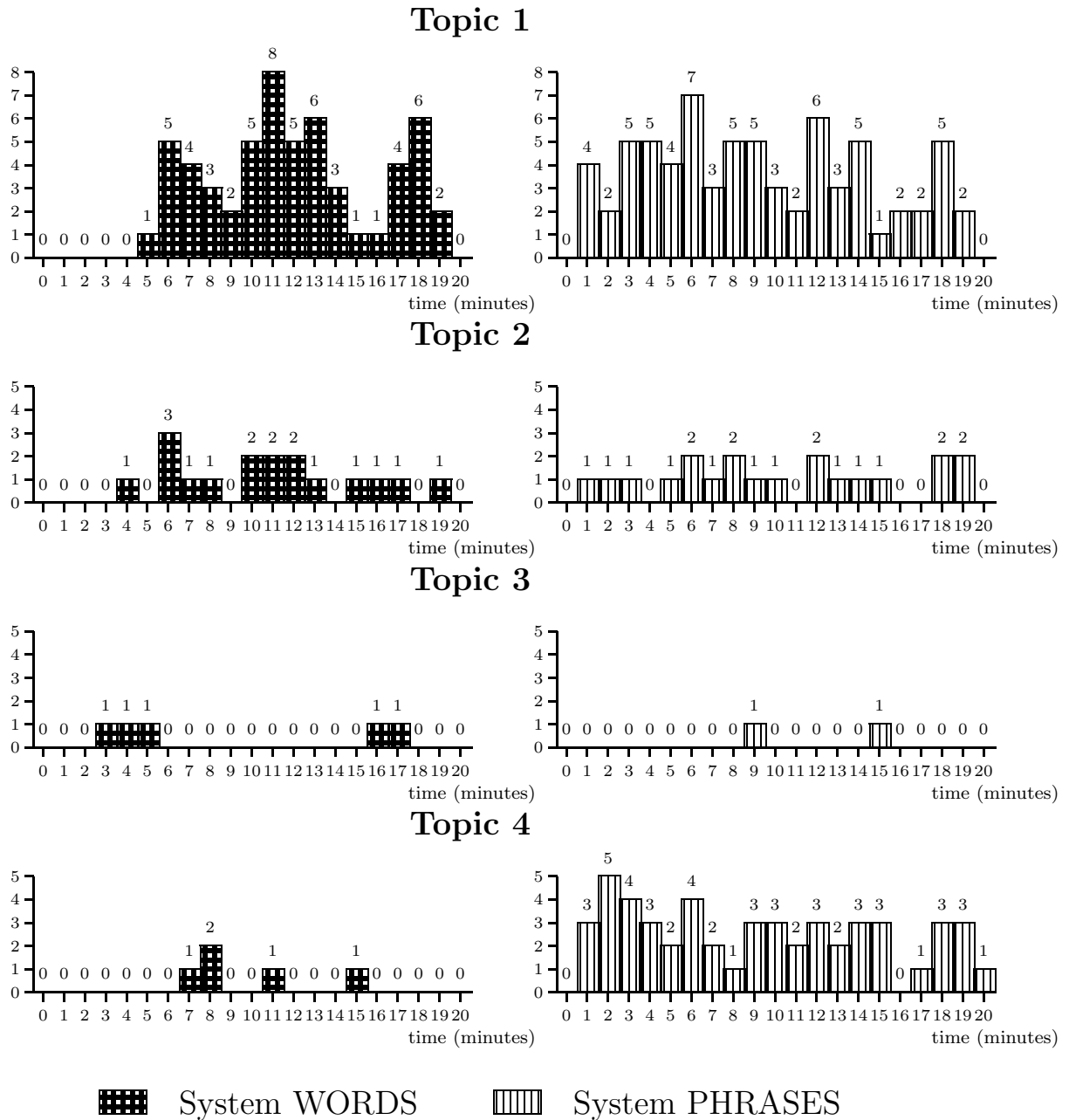


Figure 4: Distribution of number of documents selected by users across time.

The plot of document selections against time in Figure 4 provides interesting evidence about searching behavior: Searchers begin selecting documents much faster with the PHRASES system (8 selections made in minute one) than with the WORDS system (the first selection is made in minute 3). The obvious explanation is that initial query formulation is very simple in the PHRASES system (select a few phrases in the native language), and time consuming in the WORDS system (examining many foreign-language candidate translations per term and selecting them using inverse dictionary evidence).

3.2.2 Evolution of query refinements

In Figures 5, 6 and 7, the precision of the top-20 ranked lists retrieved at every refinement step is shown for topics 1, 2 and 4 (no relevant document was found for topic 3). Some observations can be made:

- The initial precision (i.e. the precision after initial query formulation) is not higher for system WORDS, in spite of the substantially higher time spent by searchers in the first query formulation. This confirms that a good initial selection of native-language phrases can provide good initial translations of the topic terms.
- Searchers perform many more query refinements with the PHRASES system, confirming that it is easier to enhance the query using phrases selected from documents.
- Searchers obtain occasional precision figures of 1, .95, .90, etc. using the PHRASES system, while the highest precision obtained with WORDS is .75 for topic 1, searcher 1.

Overall, the additional quantitative data also supports our initial hypothesis.

3.2.3 Analysis of questionnaires

The answers supplied by the eight searchers strongly support our hypothesis. All of them stated that the PHRASES system was easier to learn, easier to use and better overall. They appreciated both the ability of selecting phrases rather than individual terms, and most of them added that it was much better not to see English terms at any moment. A general claim was that the dictionary had too many acceptations for each term.

3.2.4 Observational study

The careful observation of searchers' behavior is in agreement with the above results. Some points are worth commenting:

- Users get discouraged with terms that have a lot of alternative translations in the WORDS system. Even if the term is important for the topic, they try to avoid them.
- Selecting foreign-language terms is perceived as a hard task; when no relevant documents are found after a few iterations, users get discouraged with the WORDS system.
- The refinement loop works well for the PHRASES system once relevant documents begin to appear. However, if relevant documents do not appear soon, the initial query refinements are not obvious and both systems are equally hard.
- The automatic translation of phrases may be harmful when the aligned equivalent is incorrect. This is the case of “búsqueda de tesoros”, which does not receive a correct translation (“treasure hunting”) and it is the most important concept for Topic 2. The problem is that users do not detect that the translation is incorrect; they simply think that there is no match in the collection for such concept.
- The difficulty of topic 3 (campaigns against racism in Europe) comes from the fact that the LA Times collection does not refer to any of such campaigns as generically “European”, and the overwhelming majority of documents about racism are US-centered.

4 Conclusions

We have obtained multiple evidence (quantitative data, user opinions and observational study) that a phrase-based approach to cross-language query formulation and refinement, without user-assisted translation, can be easier to use and more effective than assisted term by term translation.

Of course, this is not an absolute conclusion, if only because our reference system offered only crude help for term-by-term translation (inverse translations using a bilingual dictionary). Probably a more sophisticated translation assistance would stretch the differences between approaches. But we believe that a valid conclusion, in any case, is that language barriers are perceived as a strong impediment by users, and it is worth studying strategies of Cross-Language Search Assistance keeping a monolingual perspective from the user.

References

- [1] G. Erbach, Günter Neumann, and Hans Uszkoreit. Mulinex: Multilingual indexing, navigation and editing extensions for the world-wide web. In *AAAI Symposium on Cross-Language Text and Speech Retrieval*, 1997.
- [2] J. Gonzalo and D. Oard. The clef 2002 interactive track. In *Proceedings CLEF 2002*, 2002.
- [3] F. López-Ostenero, J. Gonzalo, A. Peñas, and F. Verdejo. Noun-phrase translations for cross-language document selection. In *Proceedings of CLEF 2001*, 2001.
- [4] W. Ogden, J. Cowie, M. Davis, E. Ludovic, S. Nirenburg, H. Molina-Salgado, and N. Sharples. Keizai: An interactive cross-language text retrieval system. In *Proceeding of the MT SUMMIT VII Workshop on Machine Translation for Cross Language Information Retrieval*, 1999.

Topic 1: Iterative rankings

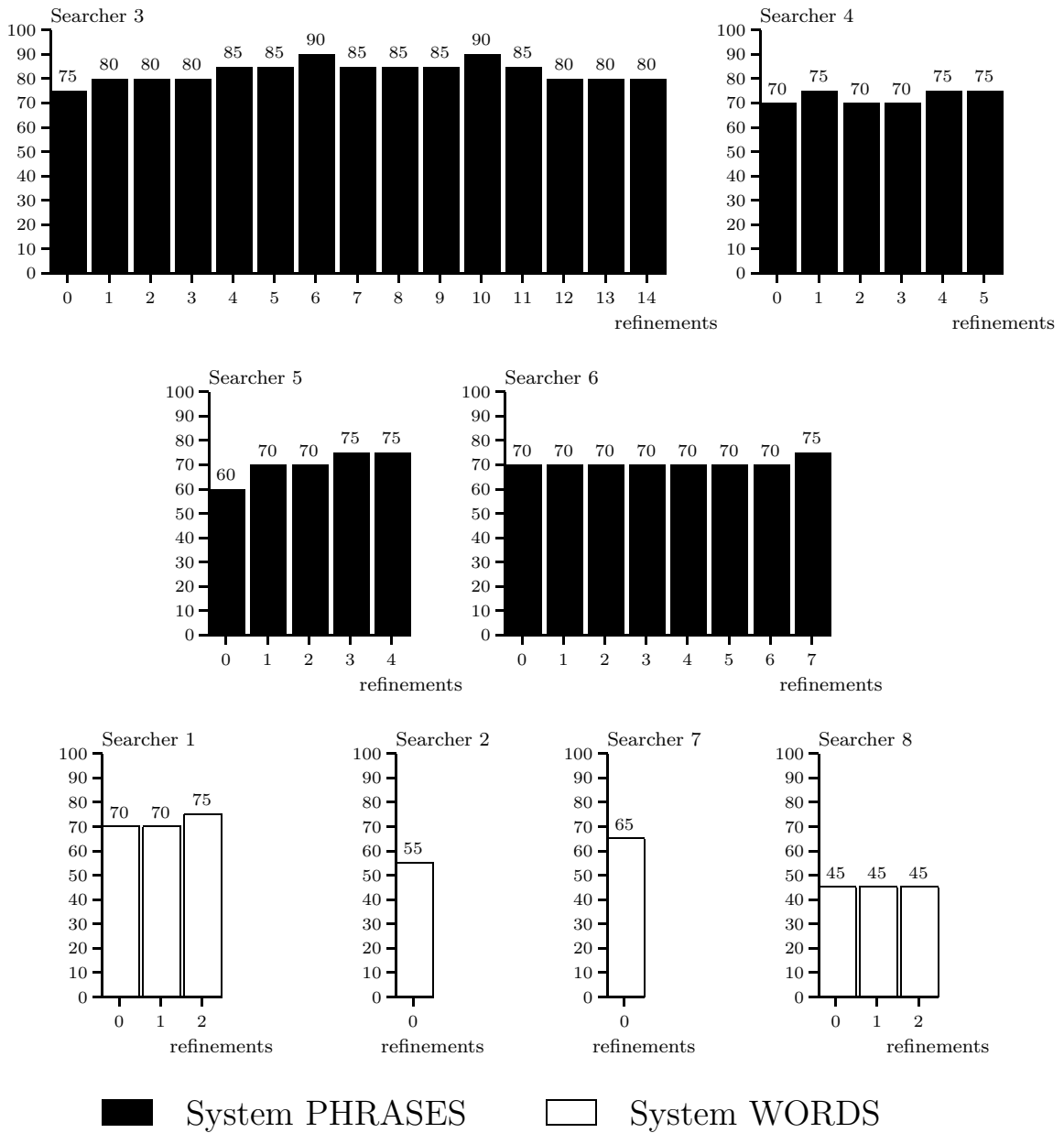


Figure 5: Precision across iterative refinements for Topic 1

Topic 2: iterative rankings

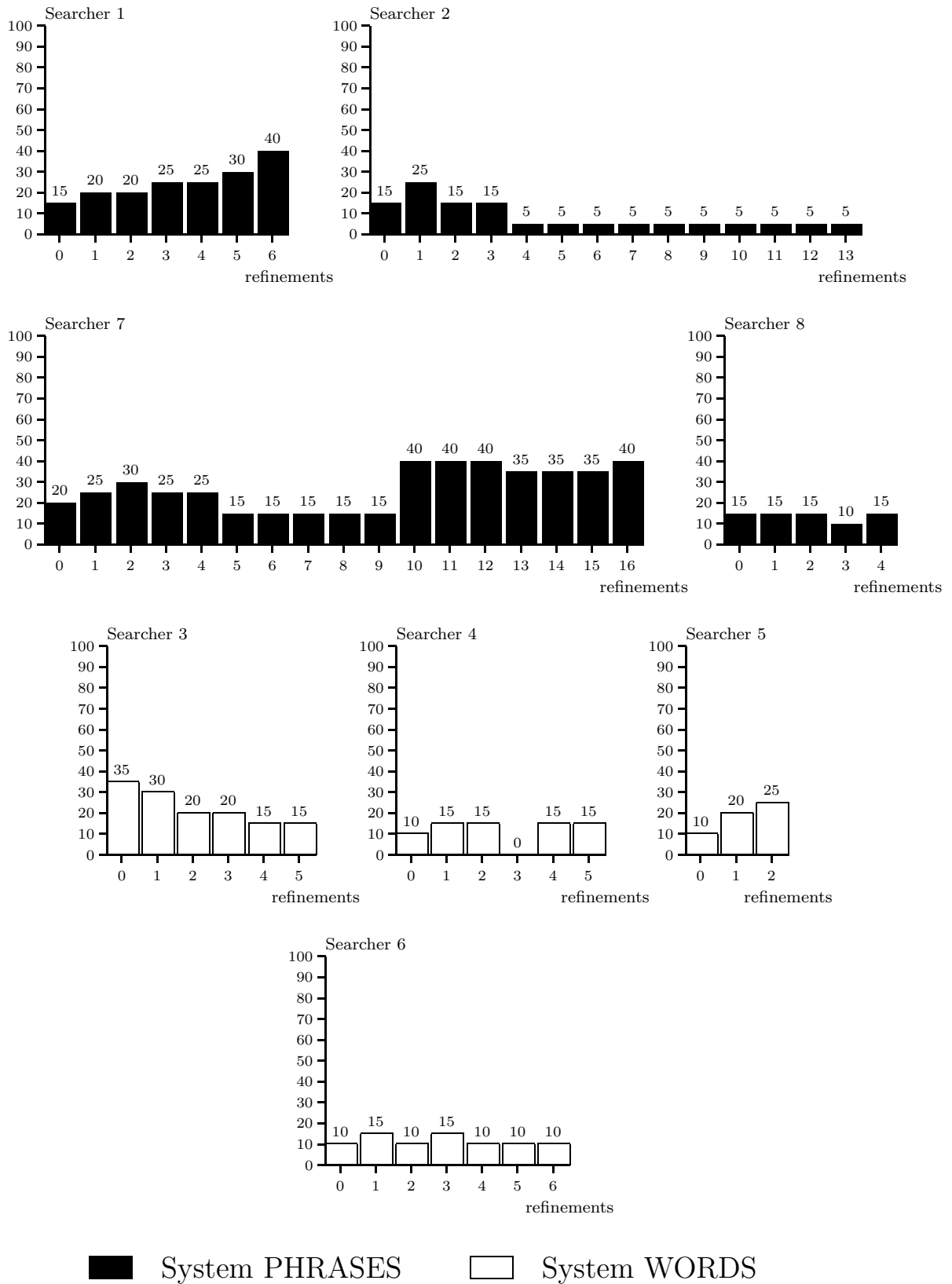


Figure 6: Precision across iterative refinements for Topic 2

Topic 4: Iterative rankings

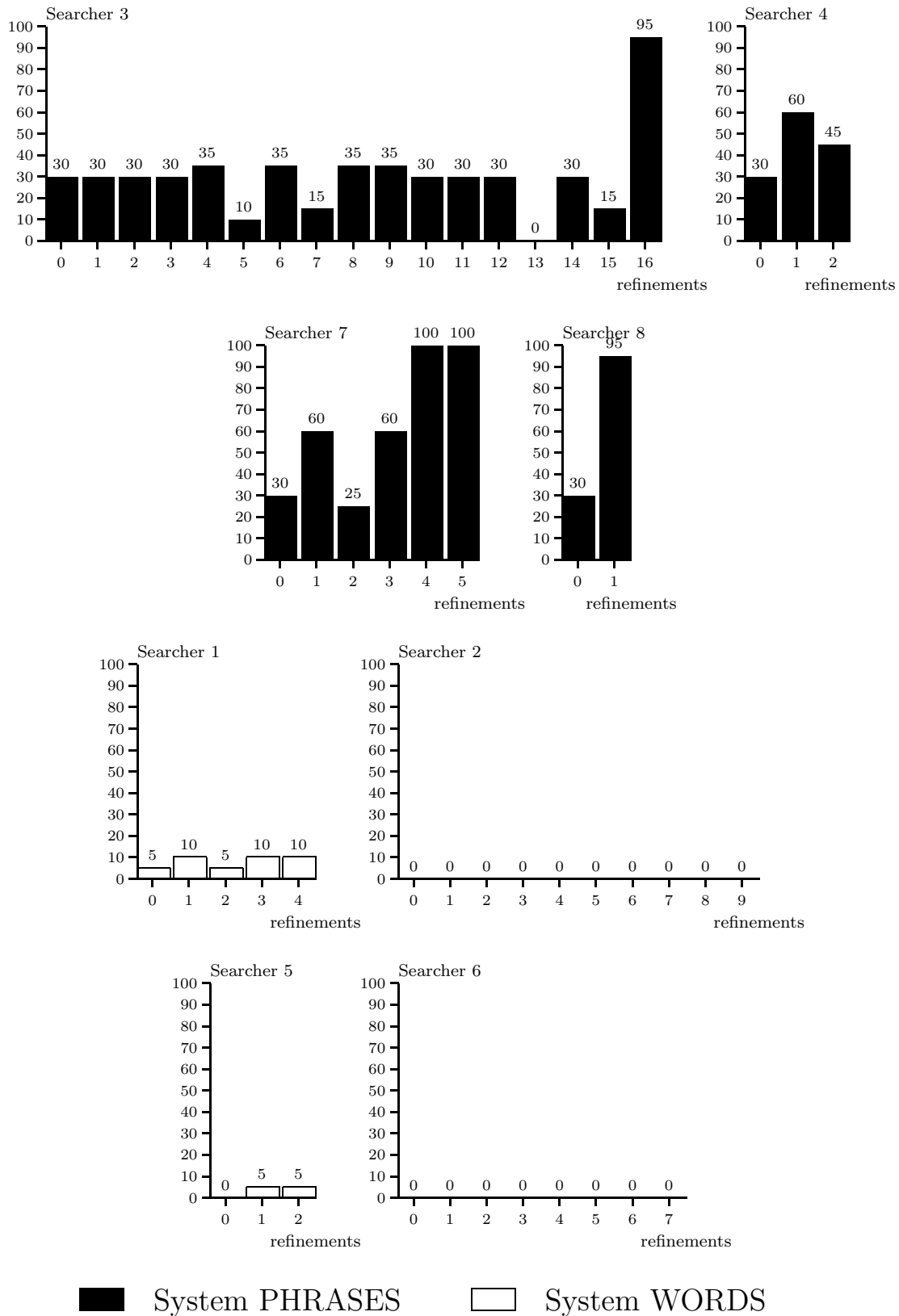


Figure 7: Precision across iterative refinements for Topic 4