Towards a Topic Complexity Measure for Cross-Language Image Retrieval

Michael Grubinger[1], Clement Leung[1], Paul Clough[2]

[1] School of Computer Science and Mathematics, Victoria University, Melbourne, Australia
michael.grubinger@research.vu.edu.au, clement@matilda.vu.edu.au
[2] Department of Information Studies, Sheffield University, Sheffield, UK
p.d.clough@sheffield.ac.uk

Abstract. Selecting suitable topics in order to assess system effectiveness is a crucial part of any benchmark, particularly those for retrieval systems. This includes establishing a range of example search requests (or topics) in order to test various aspects of the retrieval systems under evaluation. In order to assist with selecting topics, we present a measure of topic complexity for cross-language image retrieval. This measure has enabled us to ground the topic generation process within a methodical and reliable framework for ImageCLEF 2005. This document describes such a measure for topic complexity, providing concrete examples for every aspect of topic complexity and an analysis of topics used in the ImageCLEF 2003, 2004 and 2005 ad-hoc task.

1 Introduction

Benchmarks for image retrieval consist of four main elements: a collection of still natural images like [1] or [2]; a representative set of search requests (called queries or topics); a recommended set of performance measures carried out on ground truths associated with topics [3], [4]; and benchmarking events like [5] and [6] that attract participants to make use of the benchmark.

The topic selection process is a very important part of any benchmarking event. In order to produce realistic results, the topics should not only be representative of the (image) collection, but also reflect realistic user interests/needs [7]. This is achieved by generating the topics against certain dimensions, including the estimated number of relevant images for each topic, the variation of task parameters to test different translation problems, its scope (e.g. broad or narrow, general or specific), and the difficulty of the topic (topic complexity).

Hence, as the types of search request issued by users of visual information systems will vary in difficulty (or complexity), a dimension of complexity with respect to linguistic complexity for translation would help to set the context. Thus, there is a need for a measure of topic complexity that expresses the level of difficulty for retrieval systems to return relevant images in order to ground the topic generation process within a methodical and reliable framework.

As image retrieval algorithms improve, it is necessary to increase the average complexity level of topics each year in order to maintain the challenge for returning

participants. However, if topics are too difficult for current techniques the results are not particularly meaningful. Furthermore, it may prove difficult for new participants to obtain good results and prevent them from presenting results and taking part in comparative evaluations (like ImageCLEF). Providing a good variation in topic complexity is therefore very important as it allows both the organizers (and participants) to observe retrieval effectiveness with respect to complexity level.

Quantification of task difficulty is not a new concept; on the contrary, it has been applied to many areas including information retrieval [8], machine learning [9], parsing and grammatical formalisms [10], and language learning in general [11]. More recent papers include the discussion of syntactic complexity in multimedia information retrieval [12] and a measure of semantic complexity for natural language systems [13]. However, none of this research deals with the definition of a topic complexity measure for cross-language image retrieval.

This paper describes such a measure for topic complexity. Section 2 gives a short overview, examples for each aspect of topic complexity are given in sections 3 (nouns), 4 (verbs) and 5 (adjuncts). Section 6 classifies and analyses the topics used at the ImageCLEF ad-hoc tasks from 2003 to 2005. Finally, section 7 outlines further improvement of the complexity measure and other future work.

2 Overview of a Measure for Topic Complexity

The first version of the proposed scale for topic complexity starts at 0 and is unlimited as far as query difficulty is concerned. Expressed as a positive integer, the higher the value c the higher is the topic complexity.

$$0 \leq c < \infty \tag{1}$$

A value of zero implies that no translation is necessary and a simple keyword search would suffice for effective retrieval. An example for such a topic would be "David Beckham, 2005" as *David Beckham* is the same in every language[1], and so is the number *2005*.

Each of the following topic elements adds one point to topic complexity value for (cross-language) retrieval of images with complex image contents using text-based search requests:

- nouns (used as subject, direct object, indirect object or in other cases, for example genitive)
- qualifying attributes of nouns (adjectives)
- noun cardinality (grammatical number)
- verbs and qualifying attributes of verbs (adverbs)
- time, place, manner and reason adjuncts

[1] We consider only languages that use some sort of alphabet (Latin, Cyrillic, Greek, etc.) and exclude sign-based languages like Chinese etc.

In cross-language image retrieval, points are just added to the complexity level if a translation for that specific topic part is necessary (see examples hereinafter). No points are added for:

- meta-data like authors/photographers or the date of the picture
- any other part of the sentence that does not require any translation

Complexity points are accumulative: each of the elements can occur more than once and thus add more than one point (e.g. a topic can easily have two adjectives, like "*traditional Scottish* dancers"). However, logical OR constructs do not increase the complexity level if they could be expressed differently (for example: *boys or girls* is the same as *children*). The next three sections describe each of the topic elements in detail and provide example complexity scores.

3 Nouns

A word or phrase that refers to a person, place, thing, event, substance or quality is referred to as a noun (or noun substantive). Nouns can be classified in *concrete nouns* and *abstract nouns*. Concrete nouns refer to definite objects (e.g. racket, ball), whereas abstract nouns refer to ideas or concepts (e.g. fairness, freedom). In cross-language image retrieval, just concrete nouns should be used. Further, nouns can be *proper nouns* (e.g. "Michael"), *common nouns* (e.g. "boy"), or *collective nouns* (e.g. "team") and sees the use of all three types.

In topics, nouns can occur in several different cases: as *subjects* (performers of action), *direct* and *indirect objects* (recipients of action) and in the *genitive case* (indicates possession). Nouns in topics can further be described by the use of *adjectives* and have certain *cardinality*.

3.1 Subjects (Nominative Case).

The subject of a verb is the argument which generally refers to the origin of the action. In languages where a passive voice exists, the subject of a passive verb may be the target or result of the action. Passive voice should not be used in topic sentences (see also 4.1).

Each noun used as a subject (e.g. in the nominative case) increments the topic complexity level by one point.

Turtle eating leaf.
Tortuga comiendo hoja.
Schildkröte frisst Blatt.

Topic Complexity c = 3
(*Subject***,** verb, direct object)

Fig. 1. Topic complexity example for subjects. Note that articles are omitted in all examples as they are usually omitted in typical user search requests too.

3.2 Direct Objects (Accusative Case):

Objects represent the target of the verb's action. In many languages, the accusative case of a noun is, generally, the case used to mark the direct object of a verb. The accusative case exists (or existed once) in all the Indo-European languages (including Latin, Sanskrit, Greek, German, Russian), in the Finno-Ugric languages, and in Semitic languages (such as Arabic). In modern English, which lacks declension in its nouns, objects are marked by their position in the sentence or using appositions (like "to" in "I gave a book *to* him").

Each noun used as a direct object (in the accusative case) increments the topic complexity by one point:

Man riding *bicycle*.
Hombre yendo a *bicicleta*.
Mann fährt *Fahrrad*.

Topic Complexity c = 3
(Subject, verb, *direct object*)

Fig. 2. Topic complexity example for direct objects.

3.2 Indirect Objects (Dative Case):

The dative case is a grammatical case for nouns and generally marks the indirect object of a verb. Languages that use the dative case include Czech, Dutch, German, Hungarian, Icelandic, Latin, Latvian, Lithuanian, Polish, Romanian, Russian, Serbian, Croatian, Slovak, and Slovenian. In current English usage, the indirect object of an action is sometimes expressed with a prepositional phrase of "to" or "for", though an objective pronoun can also be placed directly after the main verb and used in a dative manner, provided that the verb has a direct object as well; for example, "the soccer referee shows a red card *to David Beckham*" can also be phrased as "the soccer referee shows *David Beckham* a red card".

Each noun used as an indirect object (in the dative case) increments the topic complexity by one point:

Soccer referee showing card to *soccer player*.
Arbitro mostrando una tarjeta *al futbolista*.
Schiedsrichter zeigt *Fußballspieler* eine Karte.

Topic Complexity c = 4
(Subject, verb, direct object, *indirect object*)

Fig. 3. Topic complexity example for indirect objects

3.4 Nouns in Genitive Case (or other cases)

The genitive case is a grammatical case that indicates a relationship, primarily one of possession, between the noun in the genitive case and another noun. In English, this relation can be expressed by the use of the preposition *of* ("Lord of the Rings") or by the possessive -'s ending ("Schindler's List"). Several languages have real genitive cases, including Arabic, Latin, Irish, Greek, German, Dutch, Russian, and Finnish.

Each noun used in the genitive case (or in any other grammatical case that was not mentioned here) increments the topic complexity by one point:

Man kissing *woman's* hand.
Hombre besando la mano *de una mujer*.
Mann küsst Hand *einer Frau*.

Topic Complexity c = 4
(Subject, verb, *genitive case*, direct object)

Fig. 4. Topic complexity example for the genitive case.

3.5 Qualifying Attributes of Nouns (Adjectives)

An adjective is a part of a sentence which modifies a noun, making its meaning more specific. Adjectives can be used in a *predicative* (the sky is *blue*) or *attributive* manner (the *blue* sky). In some languages e.g. the Germanic languages (like German, English, etc), attributive adjectives precede the noun. In other languages, e.g. the Romance languages (like Spanish), the adjective follows the noun. Some languages do not even have any adjectives, for example Chinese (all the words that are translated into English as adjectives are actually *stative verbs*).

Each adjective used in a topic sentence increases the topic complexity by one point. Just attributive adjectives should be used.

Austrian soccer referee showing *red* card to *Portuguese* soccer player.
Österreichischer Schiedsrichter zeigt *portugiesischem* Fußballspieler *die rote* Karte.
Arbitro *austríaco* mostrando la tarjeta *roja* a un futbolista *portugués*.

Topic Complexity c = 7
(*adjective*, noun, verb, *adjective*, direct object, *adjective*, indirect object)

Fig. 5. Topic complexity example for adjectives (compare to complexity of Fig. 3).

This example shows that a simple keyword search would not deliver good results anymore. Since keyword search does not associate attributes to the according nouns, it might as well return images showing: *Portuguese soccer referee shows card to red Austrian soccer player* or *Austrian Soccer player shows card to red, Portuguese soccer referee* etc.

3.6 Noun Cardinality (Grammatical Number, Numerals)

In linguistics, the *grammatical number* specifies the quantity of a noun or affects the form of a verb or other part of speech depending on the quantity of the noun to which it refers. Grammatical number is distinct from the use of *numerals* to specify the exact quantity of a noun.

Topics querying a specific number of a noun provide a special challenge and therefore is awarded one point for the topic complexity level.

Seven zebras drinking water.
Siete cebras bebiendo agua.
Sieben Zebras trinken Wasser.

Topic Complexity c = 4
(*cardinality*, subject, verb, direct object)

Fig. 6. Topic complexity example for noun cardinality.

4 Verbs

A verb is a part of a sentence that usually denotes action ("kick"), occurrence ("to shine"), or a state of being ("stand"). Depending on the language, a verb may vary in form according to many factors, including its tense, aspect, mood and voice. Verbs can further be described by the use of *adverbs*.

4.1 Topic Verbs

Topics should just use verbs that clearly describe the situation in an image (like running, jumping, painting, hitting, and so on). Verbs or composite verb groups that need some level of interpretation (e.g. finding, forgetting, trying to hit, attempting to escape, etc.) are not used (see also [14]).

Verbs should be used in active voice only as passive voice does not exist in all languages. Further, since the captions describe an action that is happening in the image (at that moment), the grammatically correct form for English is the present continuous tense (the auxiliary verb *to be* is omitted). In Spanish, the appropriate tense is "el presente continuo" (present continuous tense), whereas in the German

language, actions that are happening at the time are expressed with the "Präsens" (present tense).

Man *pushing* car in winter.
Hombre *empujando* coche en invierno.
Mann *schiebt* Auto im Winter.

Topic Complexity c = 4
(subject, *verb*, direct object, time adjunct)

Fig. 7. Topic complexity example for verbs. Auxiliary verbs (to be, estar) are omitted as they are also typically omitted in real user requests.

4.2 Qualifying Attributes of Verbs (Adverbs)

An adverb is a part of a sentence that serves to modify verbs, adjectives, other adverbs, and clauses. Each adverb used in a topic sentence increases the topic complexity level by one point:

Tennis player hitting ball *hard*.
Tenista golpeando la pelota *fuerte*.
Tennisspielerin schlägt Ball *hart*.

Topic Complexity c = 4
(subject, verb, direct object, *adverb*)

Fig. 8. Topic complexity example for adverbs.

Topics should just use adverbs that modify verbs. Adverbs that modify adjectives or other adverbs (adverbs of degree) are felt to be too subjective and should not be used for cross-language image retrieval of complex image contents. Adverbs of degree tell us about the intensity or degree of an action, an adjective or another adverb. Examples for adverbs of degree are: *almost, nearly, quite, just, too, enough, hardly, scarcely, completely, very, extremely*. For example: "Soccer player kicking ball *very* hard″ or *″extremely* tall boy playing basketball″.

One might argue that even the example given above may be very subjective and that adverbs should not increase the topic complexity level at all. However, as long as adverbs clearly influence the result set, they should be considered as a factor. In the example above, relevant images would include any tennis player driving, serving or smashing the ball or hitting the ball with topspin. Tennis players slicing the ball, playing a drop-volley or a stop-ball would not be relevant. This can clearly be seen by the technique of the tennis player. Hence, the adverb "hard" increases the topic complexity by one point.

4.3 Valency

The number of arguments that a verb takes is called its *valency*. According to its valency, a verb can be classified as:

- *Intransitive* (valency = 1): the verb only has a subject. For example: "people marching".
- *Transitive* (valency = 2): the verb has a subject and a direct object. For example: "golfers swinging their clubs".
- *Ditransitive* (valency = 3): the verb has a subject, a direct object and an indirect or secondary object. For example: "referee showing red card to soccer player".

It is possible to have verbs with valency = 0. A few of these appear in Spanish, Italian and other languages and are called *impersonal verbs*. For example: "Llueve" (Spanish) or "Piove" (Italian), which both mean "it rains".

Further, all languages are generally assumed to have a basic word order. Table 1 shows all possible word orders for the subject, verb, and object (in the order of the most common to the rarest).

Table 1: Word order catagories and examples.

Rk	Word Order	Example Languages
1	S-O-V languages	Turkish, Japanese, Korean, Latin, most Indian languages
2	S-V-O languages	English, Spanish, Italian, Kiswahili, Chinese, French
3	V-S-O languages	Arabic, Welsh, Gaelic
4	V-O-S languages	Fijian
5	O-S-V languages	Xavante
6	O-V-S languages	Guajiro, Hixkaryana, Klingon

For topics with valency 2 or higher, a simple keyword search is in some cases not sufficient any more as it can't detect grammatical relationships between search words. Searching for an image of "Boy chasing dog" or "Boy giving girl a candy" a simple keyword search would also return "Dog chasing boy" or "Girl giving boy a candy". In this case, due to the increased difficulty of actually having to distinguish between subject, direct and indirect object (which can't be done with the position of the noun since the word order can be different in many languages), the complexity level is increased by one point for verbs with a valency higher than 2.

5 Adjuncts

An adjunct is a type of adverbial illustrating the circumstances of the action. It expresses such relations as time, manner, place, and reason, i.e. it answers the questions: where (place adjuncts), when (time adjuncts), how (manner adjuncts) and why (adjuncts of reason).

5.1 Time Adjuncts

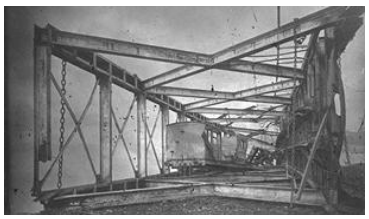
Time adjuncts indicate *when* an action happened. Topics will use prepositional phrases as time adjuncts in order to refer to time. Example:

Man riding bicycle *at night*.
Hombre yendo a bicicleta *por la noche*.
Mann fährt Fahrrad *in der Nacht*.

Topic Complexity c = 4
(subject, verb, object, *time adjunct*)

Fig. 9. Topic complexity example for time adjuncts.

If the time element does not need any form of translation, the complexity level is not increased, for example:



Tay Bridge Rail Disaster*, 1879*
Desastre del tren Tay Bridge, *1879*
Das Tay Bridge Zugsunglück, *1879*

Topic Complexity c = 1
(abstract noun)

Fig. 10. Topic complexity example for time adjuncts that do not increase the topic complexity level as no translation is necessary for 1879.

5.2 Place Adjuncts

Place adjuncts indicate the location *where* the image was taken or where the action occurred respectively. Topics will use prepositional phrases as place adjuncts. As most of the countries, cities or other places have a different name in different languages, there will always be some sort of translation involved, thus the complexity level is incremented by one.

Boat in *Northern Ireland*.
Barco en *Irlanda del Norte*.
Boot in *Nordirland*.

Topic Complexity c = 2
(noun, *place adjunct*)

Fig. 11. Topic complexity example for place adjuncts

5.3 Manner Adjuncts

Manner adjuncts further describe nouns or how actions are performed in an image. In comparison to general adverbs that modify verbs (see 3.2.2), now we just talk about *prepositional phrases* that describe *how* actions were performed.

Each manner adjunct increases the topic complexity by one point:

Woman *in bikini* leading horse *with reins*.
Mujer *en bikini* llevando un caballo *con las riendas*.
Frau *in Bikini* führt Pferd *an den Zügeln***.**

Topic Complexity c = 6
(subject, *manner adjunct*, verb, direct object, *manner adjunct*)

Fig. 12. Topic complexity example for manner adjuncts.

5.4 Adjunct of Reason

Each topic sentence can contain a reason adjunct which describes *why* actions are taken. Each reason adjunct increases the topic complexity level by one point:

Bomb damage *due to World War II.*
Bombenschäden *durch den zweiten Weltkrieg.*
Daños *por* las bombas en *la Segunda Guerra Mundial***.**

Topic Complexity c = 2
(noun, *reason adjunct*)

Fig. 13. Topic complexity example for adjuncts of reason.

Note that the English word *why* does not only refer to *reason adjuncts* (due to, because of) but also to *purpose adjuncts* (in order to). Reason adjuncts like "due to World War II" can be determined with some prior knowledge of the image, but purpose adjuncts refer to the future and can just be assumed.

Purpose adjuncts are felt to be too subjective as it involves too much interpretation of the picture and are therefore not considered for the topic complexity measure.

6 Topic Complexity at ImageCLEF

The ImageCLEF retrieval benchmark was established in 2003 with the aim of evaluating image retrieval from multilingual document collections [5][6]. This section

presents the results of the new topic complexity measure applied to the 2003, 2004 and 2005 ad-hoc ImageCLEF tasks.

6.1 Topic Complexity at ImageCLEF 2005

In the ImageCLEF 2005 ad-hoc task [15], the participants were provided with 28 topics translated into 33 different languages. Table 2 shows the analysis of the topic complexity for each of the topic titles in English.

Table 2: Topic complexity analysis for English topic titles

ID	Topic Title	Topic Analysis	c
1	aircraft on the ground	noun, place adjunct	2
2	people gathered at bandstand	noun, verb, place adjunct	3
3	dog (in) sitting (position)	noun, verb	2
4	steam ship docked	noun, noun, verb	3
5	animal statue	noun, noun	2
6	small sailing boat	adjective, noun	2
7	fishermen in boat	noun, place adjunct	2
8	building covered in snow	noun, verb, manner adjunct	3
9	horse pulling cart or carriage	noun, verb, direct object (or direct object)	3
10	sun pictures, Scotland	noun, place adjunct	2
11	Swiss mountain (scenery)	adjective, noun	2
12	postcard from Iona, Scotland	noun, place adjunct	2
13	stone viaduct with several arches	noun, manner adjunct	2
14	people at the marketplace	noun, place adjunct	2
15	golfer putting on green	noun, verb, place adjunct	3
16	waves (breaking) on beach	noun, place adjunct	2
17	man or woman reading	noun (or noun), verb	2
18	woman in white dress	noun, adjective, manner adjunct	3
19	composite postcards of Northern Ireland	adjective, noun, place adjunct, adjective	4
20	royal visit to Scotland (not Fife)	adjective, noun, place adjunct, exclusion	4
21	monument to Robert Burns	noun	1
22	building with waving flag	noun, manner adjunct, adjective	3
23	tomb inside church or cathedral	noun, place adjunct (or place adjunct)	2
24	close-up pictures of bird	noun, genitive noun	2
25	arched gateway	adjective, noun	2
26	portrait pictures of mixed-sex groups	noun, adjective, genitive noun	3
27	woman or girl carrying basket	noun (or noun), verb, direct object	3
28	colour pictures of woodland scenes around St. Andrews	adjective, noun, genitive noun, place adjunct	4

Likewise, the complexity levels have been calculated for all alphabetical languages (Romanic alphabet) with more than 10 submitted runs (e.g. European and Latin-American Spanish, Italian, German, French, Dutch, Portuguese) as same concepts are sometimes expressed with different topic complexities across various languages.

A total of 11 research groups submitted 349 runs and produced the following Mean Average Precision scores for each topic (Table 3, next page).

Table 3: Average MAP (Mean Average Precision) values for alphabetical languages with more than 10 submitted runs (with their topic complexity in parenthesis)

ID	ENG	GER	SPA – L	SPA - E	ITA	FRA	POR	NED	ALL
1	0.26 (2)	0.00 (2)	0.04 (2)	0.11 (2)	0.12 (2)	0.28 (2)	0.00 (2)	0.20 (2)	0.13 (2.00)
2	0.46 (3)	0.03 (3)	0.00 (3)	0.02 (3)	0.00 (3)	0.07 (4)	0.24 (3)	0.00 (2)	0.12 (3.00)
3	0.43 (2)	0.39 (3)	0.26(2)	0.26 (2)	0.26 (2)	0.43 (2)	0.29 (2)	0.44 (2)	0.35 (2.13)
4	0.28 (3)	0.20 (2)	0.18 (3)	0.16 (3)	0.04 (3)	0.11 (3)	0.03 (3)	0.10 (2)	0.15 (2.63)
5	0.70 (2)	0.71 (1)	0.68 (2)	0.70 (2)	0.65 (2)	0.36 (2)	0.77 (2)	0.61 (2)	0.58 (1.75)
6	0.50 (2)	0.49 (2)	0.38 (2)	0.10 (2)	0.36 (2)	0.15 (2)	0.45 (2)	0.48 (2)	0.31 (2.00)
7	0.35 (2)	0.06 (2)	0.31 (2)	0.25 (2)	0.39 (2)	0.31 (2)	0.27 (2)	0.33 (2)	0.26 (2.00)
8	0.08 (3)	0.05 (2)	0.06 (3)	0.06 (3)	0.07 (3)	0.20 (3)	0.07 (3)	0.05 (3)	0.09 (2.88)
9	0.32 (3)	0.23 (3)	0.34 (3)	0.34 (3)	0.17 (3)	0.14 (2)	0.25 (3)	0.45 (3)	0.27 (2.88)
10	0.32 (2)	0.22 (2)	0.26 (3)	0.24 (3)	0.24 (3)	0.28 (3)	0.28 (3)	0.29 (2)	0.24 (2.63)
11	0.50 (2)	0.14 (2)	0.66 (2)	0.20 (2)	0.09 (2)	0.15 (2)	0.10 (2)	0.06 (2)	0.34 (2.00)
12	0.29 (2)	0.30 (2)	0.26 (3)	0.28 (3)	0.32 (3)	0.32 (3)	0.24 (3)	0.31 (2)	0.23 (2.50)
13	0.37 (2)	0.26 (2)	0.27 (3)	0.31 (3)	0.07 (3)	0.27 (3)	0.26 (3)	0.22 (2)	0.26 (2.50)
14	0.13 (2)	0.42 (2)	0.44 (2)	0.45 (2)	0.15 (2)	0.40 (2)	0.74 (2)	0.49 (2)	0.36 (2.00)
15	0.35 (3)	0.15 (3)	0.19 (3)	0.08 (3)	0.13 (3)	0.06 (3)	0.14 (3)	0.16 (3)	0.15 (3.13)
16	0.41 (3)	0.40 (3)	0.33 (3)	0.42 (3)	0.33 (3)	0.43 (3)	0.39 (3)	0.04 (2)	0.30 (2.75)
17	0.47 (2)	0.46 (2)	0.36 (2)	0.07 (2)	0.33 (2)	0.47 (2)	0.55 (2)	0.46 (2)	0.37 (2.00)
18	0.08 (3)	0.08 (3)	0.08 (3)	0.08 (3)	0.04 (3)	0.09 (3)	0.04 (2)	0.11 (3)	0.08 (2.88)
19	0.22 (4)	0.00 (4)	0.00 (4)	0.00 (4)	0.00 (4)	0.00 (4)	0.00 (4)	0.03 (4)	0.05 (4.00)
20	0.06 (4)	0.03 (4)	0.03 (4)	0.03 (4)	0.04 (4)	0.07 (4)	0.05 (4)	0.08 (4)	0.07 (4.00)
21	0.48 (1)	0.44 (1)	0.46 (1)	0.48 (1)	0.46 (1)	0.55 (1)	0.37 (1)	0.43 (1)	0.39 (1.00)
22	0.32 (3)	0.43 (3)	0.39 (3)	0.39 (3)	0.34 (3)	0.29 (3)	0.21 (3)	0.43 (3)	0.36 (3.00)
23	0.48 (2)	0.34 (2)	0.33 (2)	0.06 (2)	0.02 (2)	0.08 (2)	0.26 (2)	0.54 (2)	0.22 (2.00)
24	0.22 (2)	0.25 (2)	0.15 (2)	0.12 (2)	0.16 (2)	0.17 (2)	0.23 (2)	0.26 (2)	0.19 (2.00)
25	0.45 (2)	0.13 (2)	0.07 (2)	0.11 (2)	0.03 (2)	0.38 (2)	0.22 (2)	0.06 (2)	0.19 (2.00)
26	0.53 (3)	0.36 (3)	0.22 (3)	0.15 (3)	0.08 (3)	0.29 (2)	0.10 (3)	0.37 (3)	0.25 (2.88)
27	0.35 (3)	0.28 (3)	0.14 (3)	0.15 (3)	0.21 (3)	0.29 (3)	0.08 (3)	0.33 (3)	0.22 (3.00)
28	0.13 (4)	0.13 (3)	0.12 (4)	0.10 (4)	0.10 (3)	0.12 (4)	0.09 (4)	0.15 (3)	0.11 (3.63)

In order to establish the existence of a relation between the level of complexity and results obtained from ImageCLEF submissions, the correlation coefficient $\rho_{x,y}$ is calculated for each of the languages, using the following formula:

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y} \tag{2}$$

where

$$-1 \leq \rho_{XY} \leq 1 \tag{3}$$

and

$$Cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_X)(y_i - \mu_Y) \tag{4}$$

where X corresponds to the complexity levels for each topic and Y to their respective results.

Figure 14 shows that a strong negative correlation exists between the level of topic complexity and MAP of submitted ImageCLEF results (the higher the topic complexity score, the lower the MAP score).

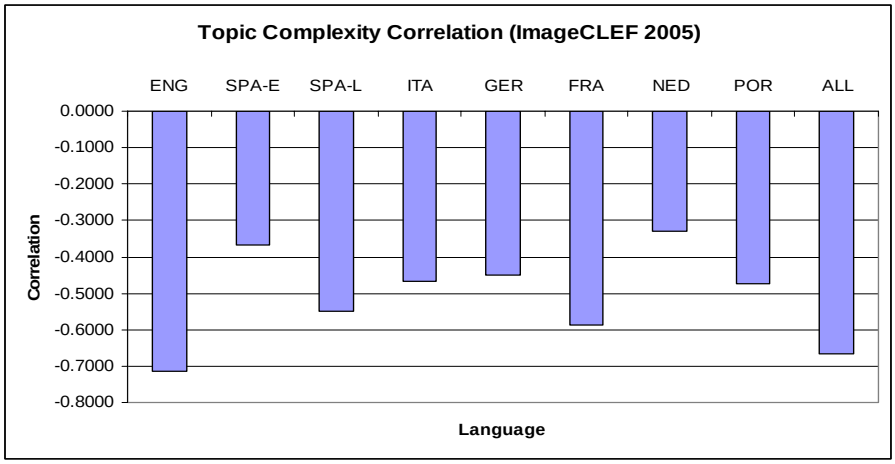


Fig. 14. Correlation between topic complexity score and MAP for ImageCLEF 2005 submissions.

6.2 Topic Complexity at ImageCLEF 2004

In ImageCLEF 2004 [6], twelve participating groups submitted 190 runs to the ad-hoc task. Similar to results in section 6.1, the levels of topic complexity were calculated for all topics and compared with the average MAP results for languages with more than 10 submissions.

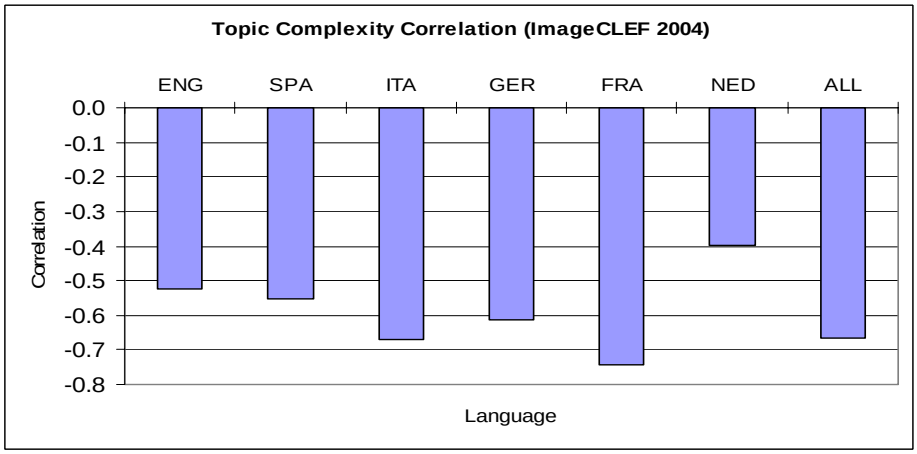


Fig. 15. Topic complexity correlation for ImageCLEF 2004

Figures 15 shows, again, a strong negative correlation. The correlation factor $\rho_{x,y}$ is always stronger than -0.4, reaches more than -0.6 for Italian and German and even more than -0.7 for French.

6.3 Topic Complexity at ImageCLEF 2003

In ImageCLEF 2003 [5], four participating groups submitted 45 runs to the ad-hoc task. Similar to results in sections 6.1 and 6.2, the levels of topic complexity were calculated for all topics and compared with the average MAP results for languages with more than 5 submissions.

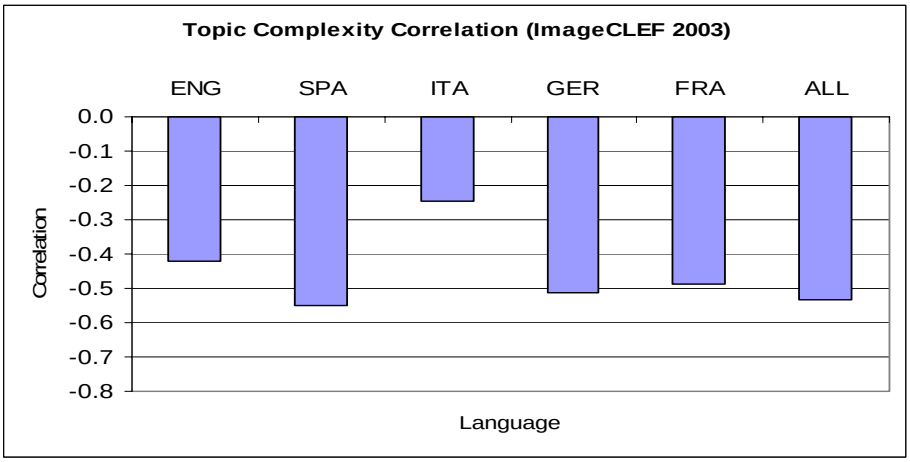


Fig. 16. Topic complexity correlation for ImageCLEF 2003

The results shown in Figure 16 show a very strong negative correlation again. Like in 2004 and 2005, the correlation factor $\rho_{x,y}$ is always stronger than -0.4 (except for Italian which is due to a couple of translation problems that produced surprising results).

7 Conclusion and Future Work

In this paper, we present a measure for the degree of topic complexity for search requests of cross-language image retrieval. Establishing such a measure is beneficial when creating benchmarks such as ImageCLEF in that it is possible to categorise results according to a level of complexity for individual topics. This can help explain results obtained when using the benchmark and provide some kind of control and reasoning over topic generation.

Examples illustrating various aspects of the linguistic structure of the complexity measure and motivating its creation have been presented. Comparing the level of complexity for topics created in ImageCLEF 2003 to 2005 for the ad-hoc task with MAP scores from submitted runs by participating groups have shown a strong nega-

tive correlation indicating that more linguistically complex topics result in much lower MAP scores due to the requirement of more complex translation approaches.

Future work will involve the improvement and refinement of the complexity measure and further verification by analysing results from the 2006 ImageCLEF ad-hoc task.

References

1. Clough, P., Sanderson, M., Reid, N.: The Eurovision St Andrews Photographic Collection (ESTA), Image CLEF Report, University of Sheffield, UK (February 2003).
2. Grubinger, M., Leung, C.: A Benchmark for Performance Calibration in Visual Information Search, In Proceedings of 2003 Conference of Visual Information Systems, Miami, Florida, USA (September 24-26, 2003) 414 – 419.
3. Leung, C., Ip, H.: Benchmarking for Content Based Visual Information Search. In Laurini ed. Fourth International Conference on Visual Information Systems (VISUAL'2000), Lecture Notes in Computer Science, Springer Verlag, Lyon, France (November 2000) 442 – 456.
4. Müller, H., Müller, W., Squire, D., Marchand-Millet, S., Pun, T.: Performance Evaluation in Content Based Image Retrieval: Overview and Proposals. Pattern Recognition Letters, 22(5), (April 2001), 563 – 601.
5. Clough, P., Sanderson, M.: The CLEF 2003 cross language image retrieval task. In Proceedings of the Cross Language Evaluation Forum (CLEF) 2003, Springer Verlag, (2004).
6. Clough, P., Müller, H., Sanderson, M.: Overview of the CLEF cross-language image retrieval track (ImageCLEF) 2004. In C. Peters, P.D. Clough, G. J. F. Jones, J. Gonzalo, M. Kluck, and B. Magnini, editors, Multilingual Information Access for Text, Speech and Images: Result of the fifth CLEF evaluation campaign, Lecture Notes in Computer Science, Springer Verlag, Bath, England (2005).
7. Armitage, L., Enser, P.: Analysis of User Need in Image Archives. Journal of Information Science, (23 (4) 1997), 287 – 299.
8. Bagga, A., Biermann, A.: Analysing the complexity of a domain with respect to an information extraction task. In Proceedings of the tenth International Conference on Research on Computational Linguistics (ROCLING X), (August 1997), 175 – 194.
9. Niyogi, P.: The Informational Complexity of Learning from Examples. PhD Thesis, MIT. (1996).
10. Barton, E., Berwick, R., Ristad, E.: Computational Complexity and Natural Language. The MIT Press, Cambridge, Massachusetts (1987).
11. Ristad, E.: The Language Complexity Games. MIT Press, Cambridge, MA, (1993).
12. Flank, S.: Sentences vs. Phrases: Syntactic Complexity in Multimedia Information Retrieval. NAACL-ANLP 2000 Workshop: Syntactic and Semantic Complexity in Natural Language Processing Systems (2000).
13. Pollard, S., Biermann, A.: A Measure of Semantic Complexity for Natural Language Systems. NAACL-ANLP 2000 Workshop: Syntactic and Semantic Complexity in Natural Language Processing Systems.
14. Grubinger, M., Leung, C.: Incremental Benchmark Development and Administration, In Proceedings of 2004 Conference of Visual Information Systems, Hotel Sofitel Conference Centre, San Francisco, California, USA (September 8-10, 2004).
15. Clough, P., Desealers, T., Grubinger, M., Hersh, W., Jensen, J., Lehmann, T., Müller, H.: The 2005 Cross-Language Image Retrieval Track. In these proceedings (2005).