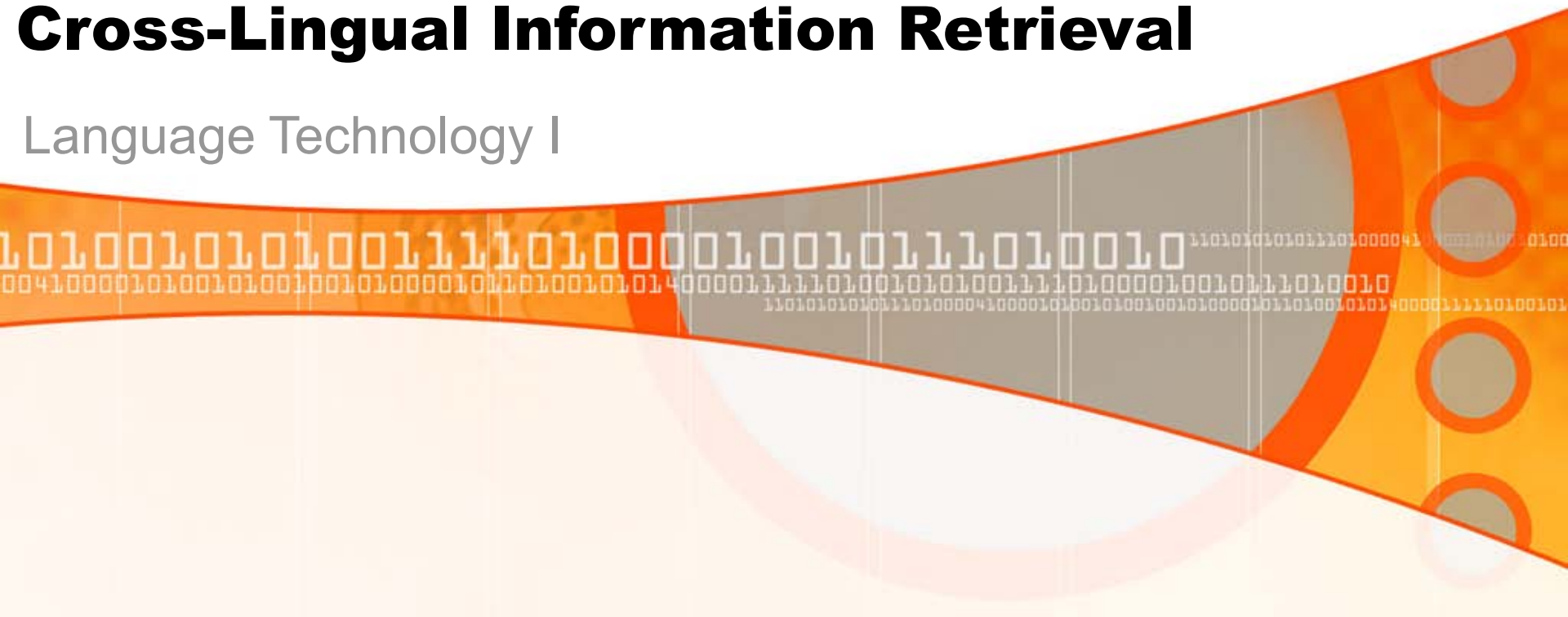
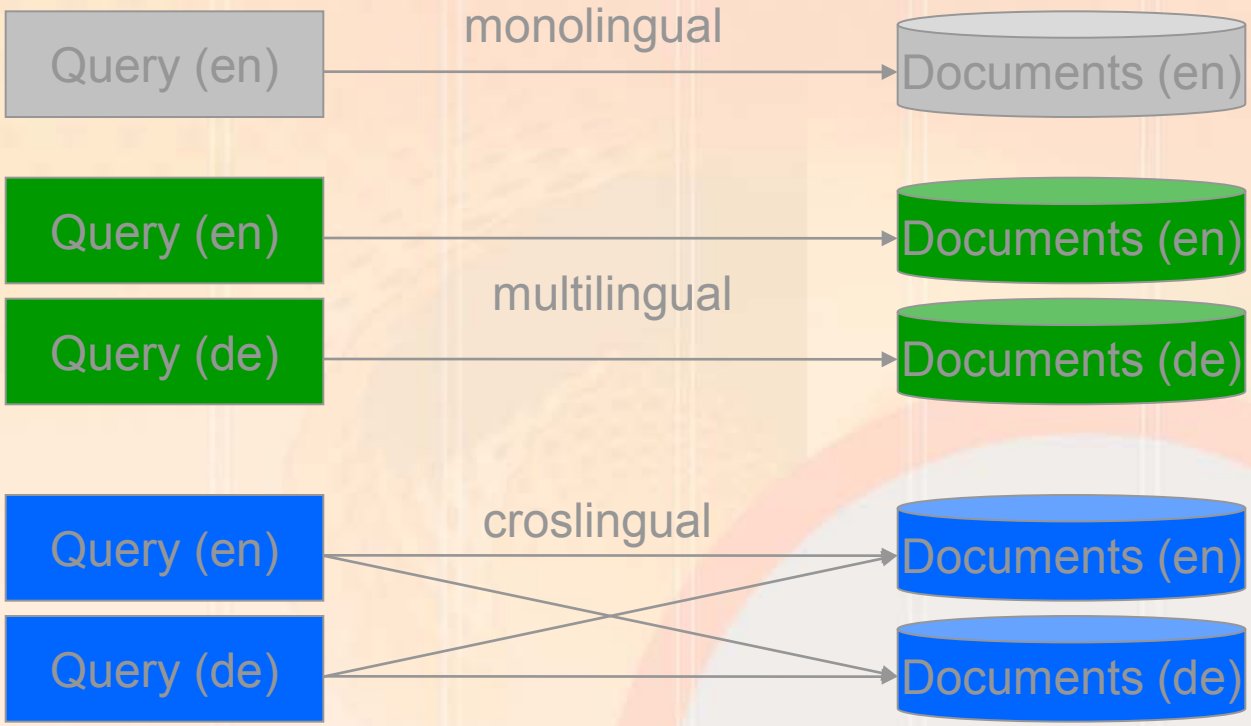
Cross-Lingual Information Retrieval

Language Technology I



Terminology

- monolingual, multilingual, cross-lingual



Use Scenarios (I)

- a user has no knowledge of a target language, i.e., she cannot search for documents in that language at all
  - *with CLIR she can make use of media data pools that are indexed with captions in that language, for example for picture pools, music databases, etc.*
  - *with CLIR she can get a pre-selection of documents that can then be passed on to a translator*

Use Scenarios (II)

- a user has only passive knowledge of a target language, i.e., she cannot actively search for documents in that language
  - *with CLIR she can make use of relevant texts*

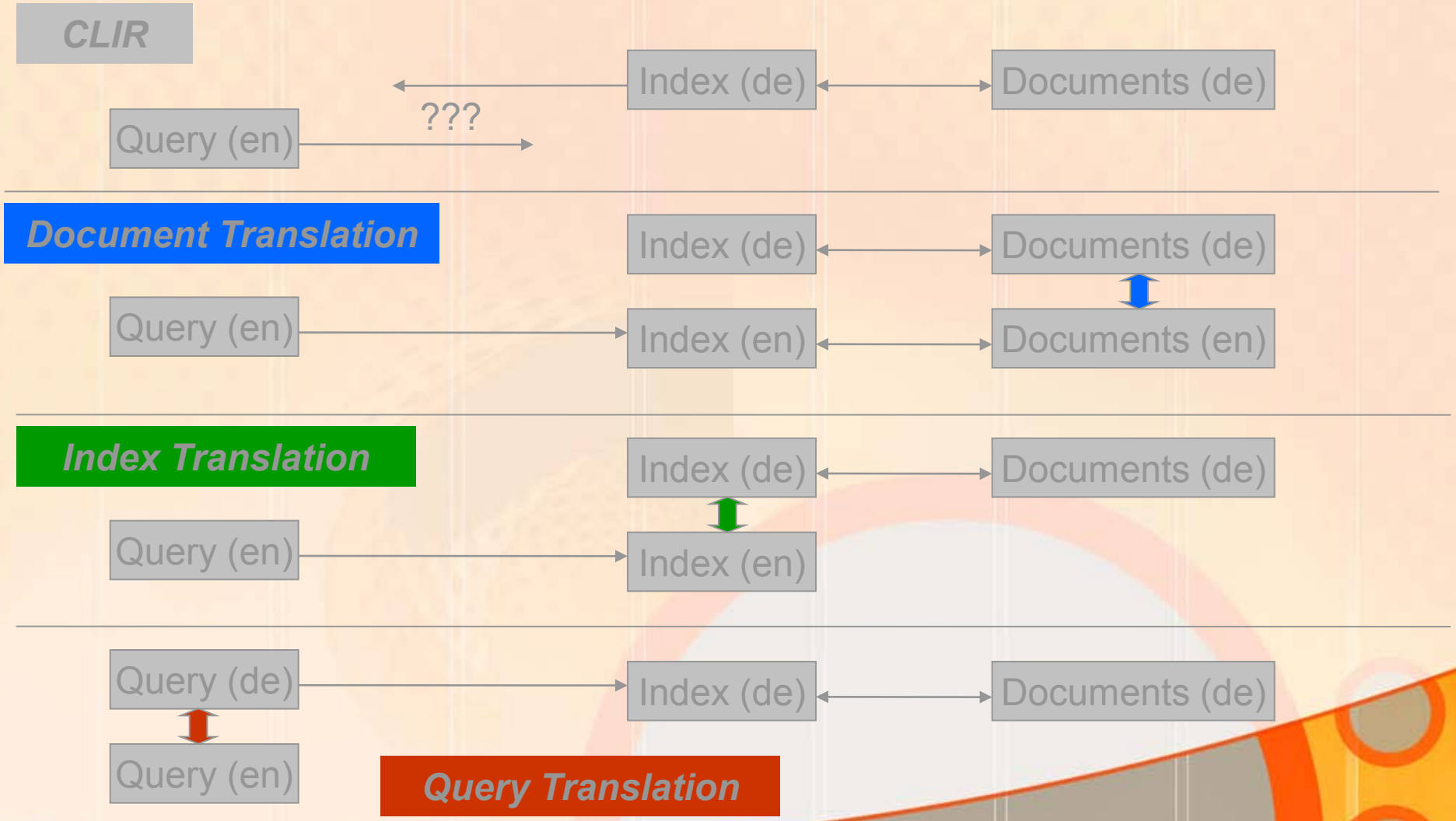
Use Scenarios (III)

- a document collection has such a large number of languages that it would be impractical to formulate a query in each of these languages
  - *with CLIR one could get relevant documents with only a search query in one of these languages*

CLIR approaches

- Machine translation:
  - *uses NLP tools like PoS-tagger, parser, morphological analyzers, etc.*
- Thesaurus-based approaches
  - *manual use of thesauri: "controlled vocabulary" systems*
  - *automatic use of thesauri: "concept retrieval" systems*
- Corpus-based methods: work with frequency analysis
  - *Implication: aboutness of the two collections should be similar*

MT Approach - Architecture



Document Translation

- Problem solved by multiplying the texts
  - *Make texts available in all languages*
  - *multilingual (= several monolingual) retrieval*
- Feasibility:
  - *Required in some applications*
    - *Patents, multilingual states (EG, Belgium, …)*
  - *Impossible in other areas (Internet)*
- Evaluation:
  - *From costly to impossible*
  - *Results depend on translation quality*
    - *translation dictionary updates invalidate search on existing document pool (->retranslate everything)*

Index Translation

- Idea:
  - *multilingual Index*
    - *Analyze query in query language, translate terms*
    - *Search with all document language index terms*
  - *(Problem of retranslation of the hits)*
- Feasibility:
  - *Not feasible*
    - *Ambiguity of index terms*
    - *Multiword terms not in index*
    - *Context dependency of translations*

Fehler: mistake, fault, error, bug
nuclear: Kern~, zentral, nuklear
power: Macht, Kraft, Strom
plant: Pflanze, Unternehmen

=> Organize the index as a special resource!

Query Translation

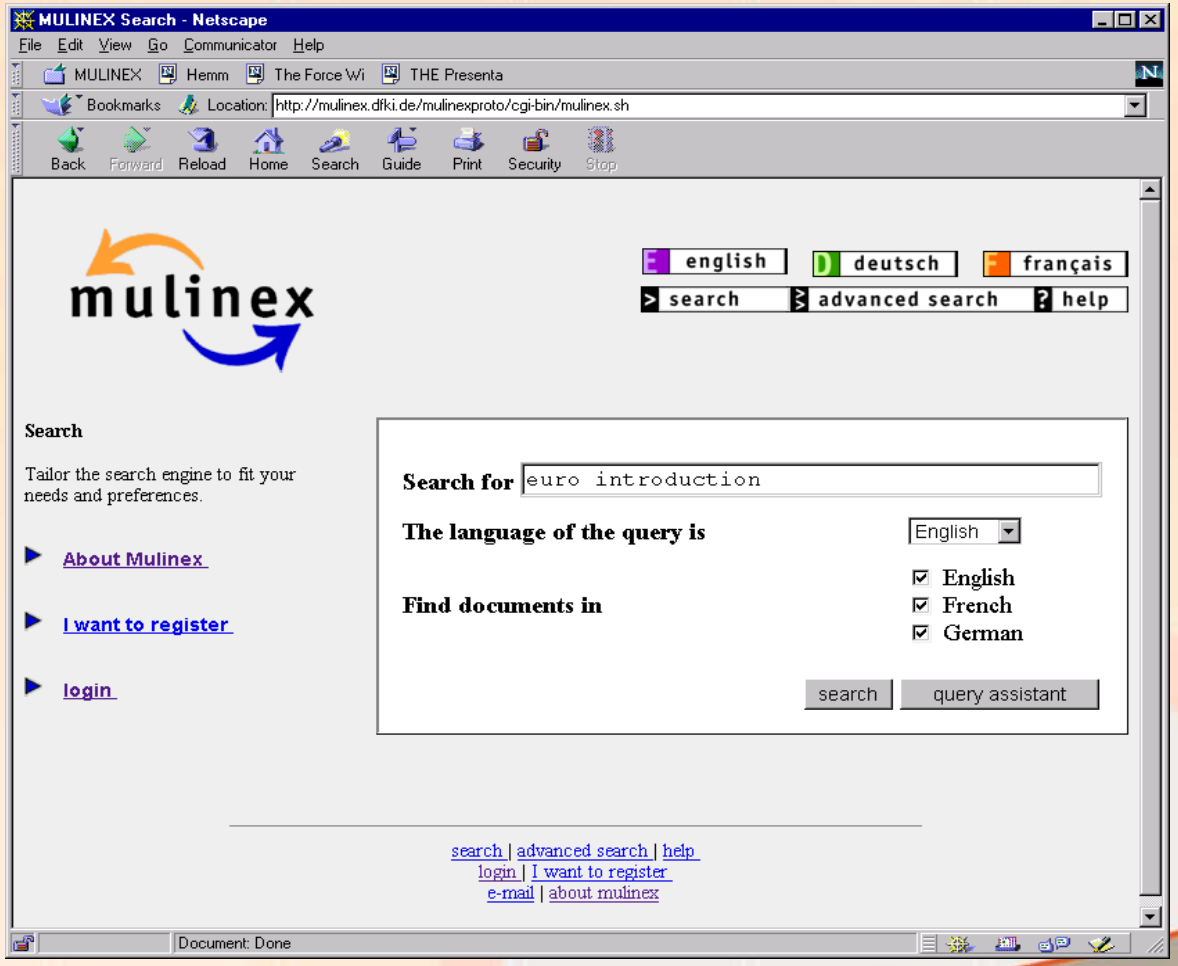
- Approach: Translation of query
  - *Analyse and translate the query terms*
  - *Search in (monolingual) Backend-System*
- Evaluation
  - *Backend database stays unchanged*
  - *Translation changes do not affect document base*
  - *Cross-lingual component as system frontend*
    - *contains multilingual linguistic resource*
    - *Which is also usable for re-translation*
    - *And can be maintained independently*
  - *Cross-linguality is transparent for the users*
  - *Fine-tuning between frontend and backend required*

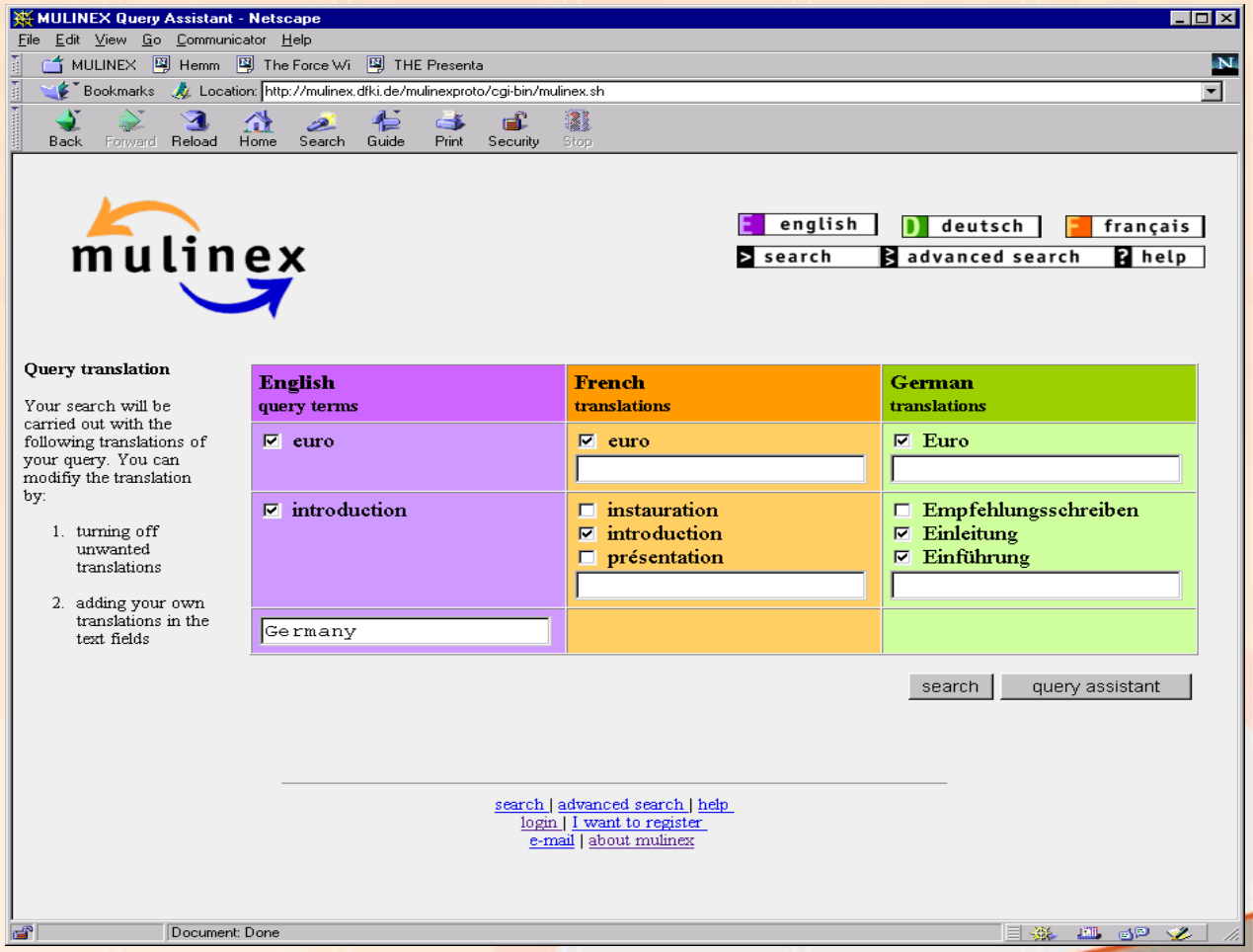
MT Approach

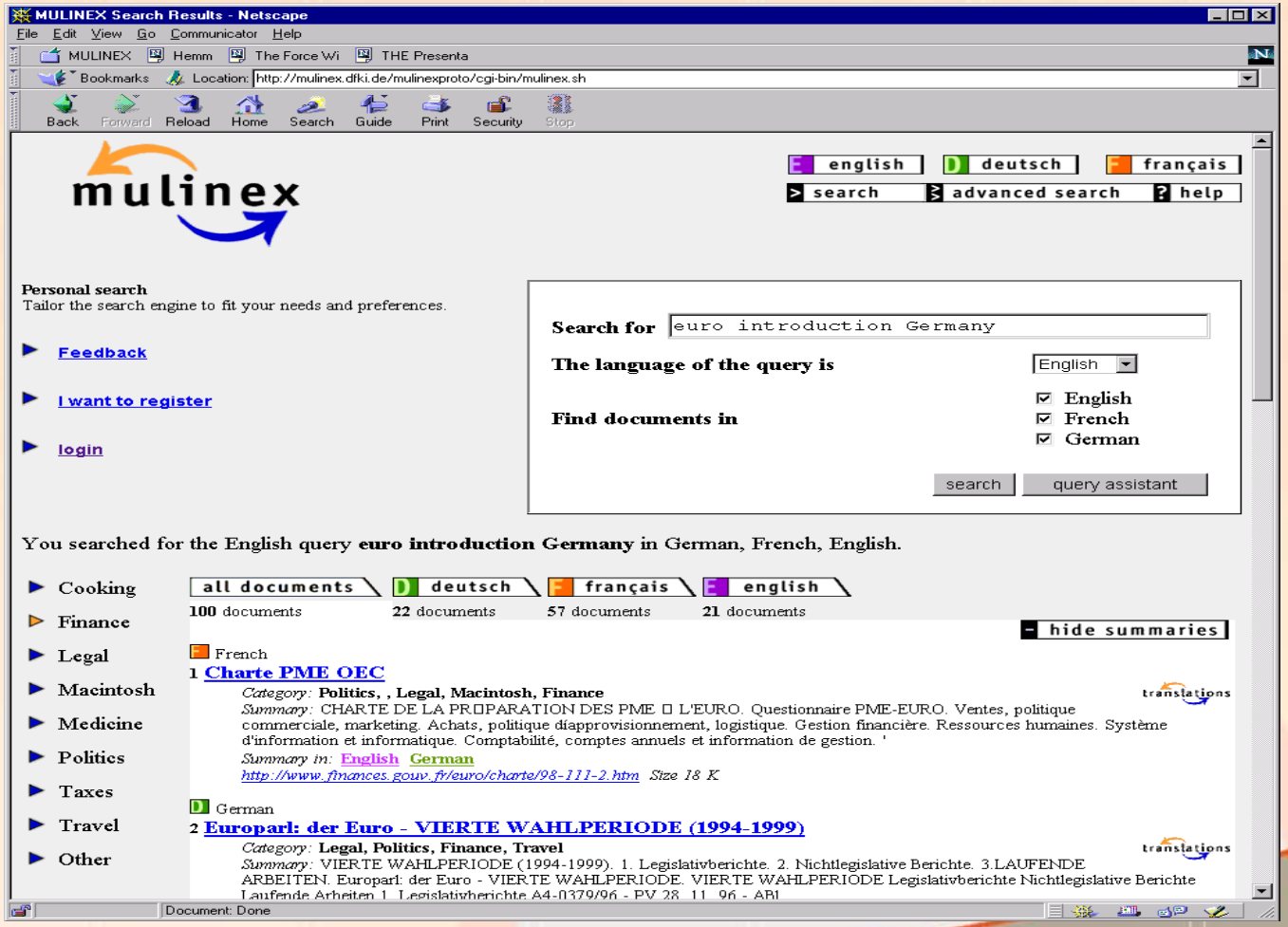
- pros:
  - *straightforward (if an MT system is available)*
  - *user can directly use the retrieved documents*
  - *documents usually have more context which allows more robust MT than for query translation*
- cons:
  - *translation of document collections may be very time consuming*
  - *offline translation of document collections may require lots of additional storage*
  - *inherits most weaknesses of MT and MT system implementations*

Thesaurus-Based Approach: “Thesauri”

- thesaurus: a resource which organizes the terminology of a domain of knowledge, i.e., an ontology for terminology
- multilingual thesauri encode
  - *usually: cross-linguistic synonymy*
  - *sometimes: hierarchical relations between terms (hyperonymy,hyponymy, etc.)*
  - *seldom: associative relations between terms*
- the thesaurus-based approach to CLIR
  - *uses multilingual thesauri*
  - *has a rather broad definition of a thesaurus*
- examples of multilingual thesauri used for CLIR:
  - *simple cross-language synonym lists*
  - *collection of concepts with attached cross-lingual information*
  - *“classic” syntax and semantics lexicons*


MULINEX Search - Netscape
File Edit View Go Communicator Help
MULINEX Hemm The Force Wi THE Presenta
Bookmarks Location: http://mulinex.dfki.de/mulinexproto/cgi-bin/mulinex.sh
Back Forward Reload Home Search Guide Print Security Stop
mulinex
english deutsch français
search advanced search help
Search
Tailor the search engine to fit your needs and preferences.
About Mulinex
I want to register
login
Search for euro introduction
The language of the query is English
Find documents in
English
French
German
search query assistant
search | advanced search | help
login | I want to register
e-mail | about mulinex
Document: Done


MULINEX Query Assistant - Netscape
File Edit View Go Communicator Help
MULINEX Hemm The Force Wi THE Presenta
Bookmarks Location: http://mulinex.dfki.de/mulinexproto/cgi-bin/mulinex.sh
Back Forward Reload Home Search Guide Print Security Stop
mulinex
english deutsch français
search advanced search help
Query translation
Your search will be carried out with the following translations of your query. You can modifiy the translation by:
1. turning off unwanted translations
2. adding your own translations in the text fields
English query terms
French translations
German translations
euro
euro
Euro
introduction
instauration
introduction
présentation
Empfehlungsschreiben
Einleitung
Einführung
Germany
search
query assistant
search | advanced search | help
login | I want to register
e-mail | about mulinex
Document: Done


MULINEX Search Results - Netscape
File Edit View Go Communicator Help
MULINEX Hemm The Force Wi THE Presenta
Bookmarks Location: http://mulinex.dfki.de/mulinexproto/cgi-bin/mulinex.sh
Back Forward Reload Home Search Guide Print Security Stop
mulinex
english deutsch français
search advanced search help
Personal search
Tailor the search engine to fit your needs and preferences.
Feedback
I want to register
login
Search for euro introduction Germany
The language of the query is English
Find documents in
English
French
German
search query assistant
You searched for the English query euro introduction Germany in German, French, English.
Cooking
Finance
Legal
Macintosh
Medicine
Politics
Taxes
Travel
Other
all documents deutsch français english
100 documents 22 documents 57 documents 21 documents
hide summaries
French
1 Charte PME OEC
Category: Politics, , Legal, Macintosh, Finance
Summary: CHARTE DE LA PR□PARATION DES PME □ L'EURO. Questionnaire PME-EURO. Ventes, politique commerciale, marketing. Achats, politique d'approvisionnement, logistique. Gestion financière. Ressources humaines. Système d'information et informatique. Comptabilité, comptes annuels et information de gestion. '
Summary in: English German
http://www.finances.gouv.fr/euro/charte/98-111-2.htm Size 18 K
translations
German
2 Europarl: der Euro - VIERTE WAHLPERIODE (1994-1999)
Category: Legal, Politics, Finance, Travel
Summary: VIERTE WAHLPERIODE (1994-1999). 1. Legislativberichte. 2. Nichtlegislative Berichte. 3.LAUFENDE ARBEITEN. Europarl: der Euro - VIERTE WAHLPERIODE. VIERTE WAHLPERIODE Legislativberichte Nichtlegislative Berichte Laufende Arbeiten 1. Legislativberichte A4-0379/96 - PV 28. 11. 96 - API
translations
Document: Done

Thesaurus-Based Approach: “Thesauri”

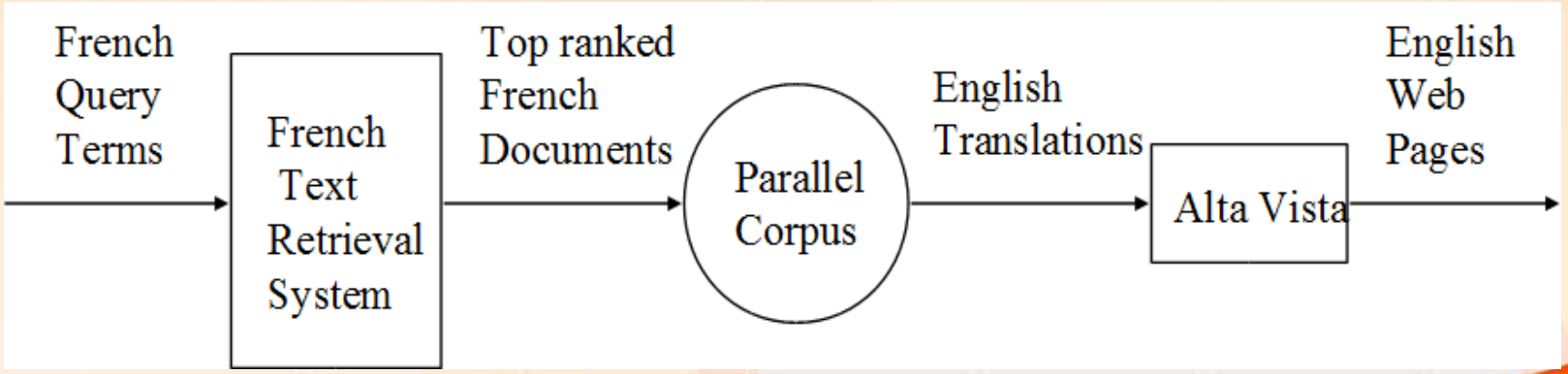
- pros:
  - *very productive, especially for skilled users*
  - *works transparently for the user*
  - *unambiguous mapping between the query and the target document*
- cons:
  - *very expensive to create good thesauri*
  - *target documents must be labeled with concepts*
  - *may be difficult to use for unexperienced users (e.g., because of the manual selection of the intended concept)*
  - *doesn’t scale*
  - *restricted to certain domains*
  - *IR queries can only be as precise as the predefined thesaurus concepts*

Corpus-Based Approach

- use of statistical information about term usage from parallel corpora
- usually based on two general retrieval principles:
  - *target documents with frequent usage of query terms are potentially more relevant than target documents with infrequent query term usage*
  - *rare query terms are more useful than query terms that are very frequent in the overall target document collection*
- pros:
  - *usage of recent terminology (as provided by the corpora) is possible*
- cons:
  - *parallel corpora needed*
  - *restricted to the domains of the parallel corpora*

Pseudo-Relevance Feedback

- Enter query terms in French
- Find top French documents in parallel corpus
- Construct a query from English translations
- Perform a monolingual free text search



Learning From Document Pairs

- Count how often each term occurs in each pair
  - *Treat each pair as a single document*

	English Terms					Spanish Terms			
	E1	E2	E3	E4	E5	S1	S2	S3	S4
Doc 1	4		2			2			1
Doc 2	8		4			4			2
Doc 3		2		2			2	1	
Doc 4		2	1				2		1
Doc 5	4				1	2		1	

Similarity based Dictionaries

- Automatically developed from aligned documents
  - *Terms E1 and E3 are used in similar ways*
    - *Terms E1 & S1 (or E3 & S4) are even more similar*
- For each term, find most similar in other language
  - *Retain only the top few (5 or so)*

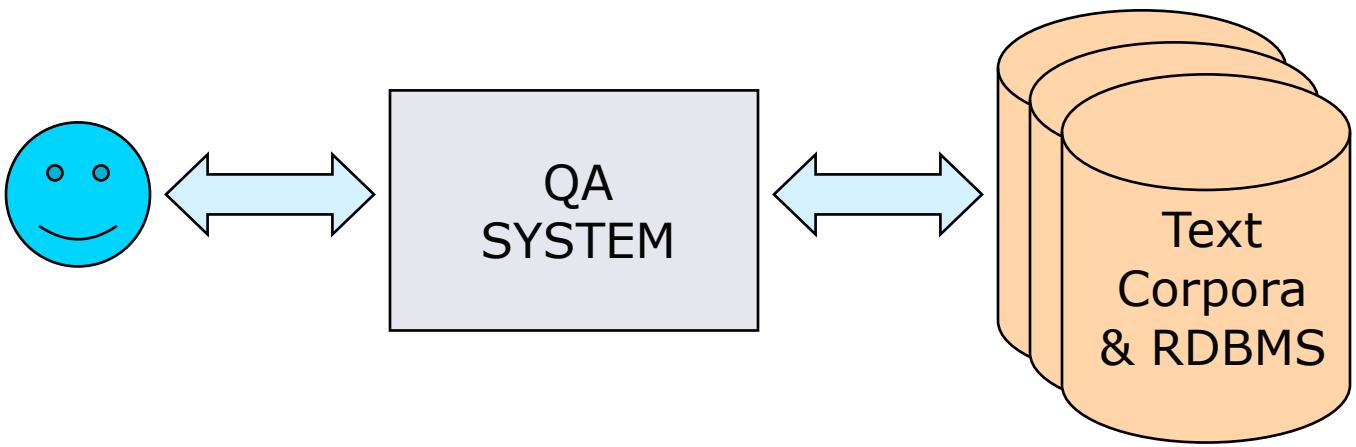
CLIR Research Community

- Text REtrieval Conference (TREC, http://trec.nist.gov/)
  - *Arabic, English, Spanish, Chinese, etc.*
  - *CLIR at TREC: http://www.glue.umd.edu/~dlrg/clir/trec2002/*
- Cross-Language Evaluation Forum (CLEF)
  - *European languages*
  - *http://www.clef-campaign.org/*
- NTCIR (NII Test Collection for IR Systems)
  - *http://research.nii.ac.jp/ntcir/index-en.html*
  - *with related workshops*
- Information Retrieval for Asian Language (IRAL)
  - *internaltional workshop*
- and quite a few others

Open-domain Cross-lingual Question Answering from Unstructured Documents

Question Answering

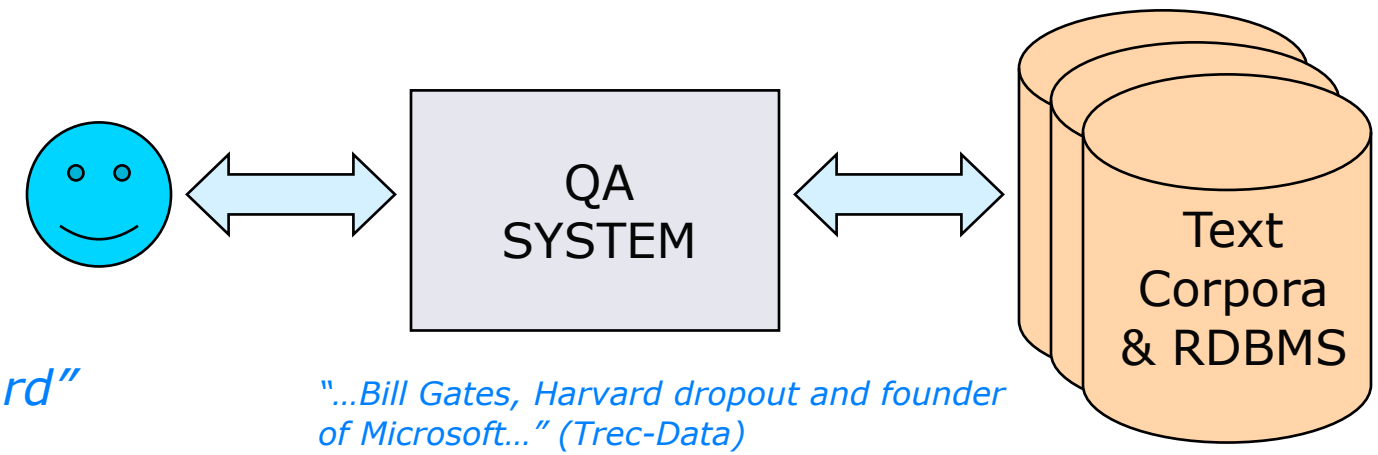
- Input: a question in NL; a set of text and database resources
- Output: a set of possible answers drawn from the resources



Question Answering

- Input: a question in NL; a set of text and database resources
- Output: a set of possible answers drawn from the resources

"Where did Bill Gates go to college?"



"Harvard"

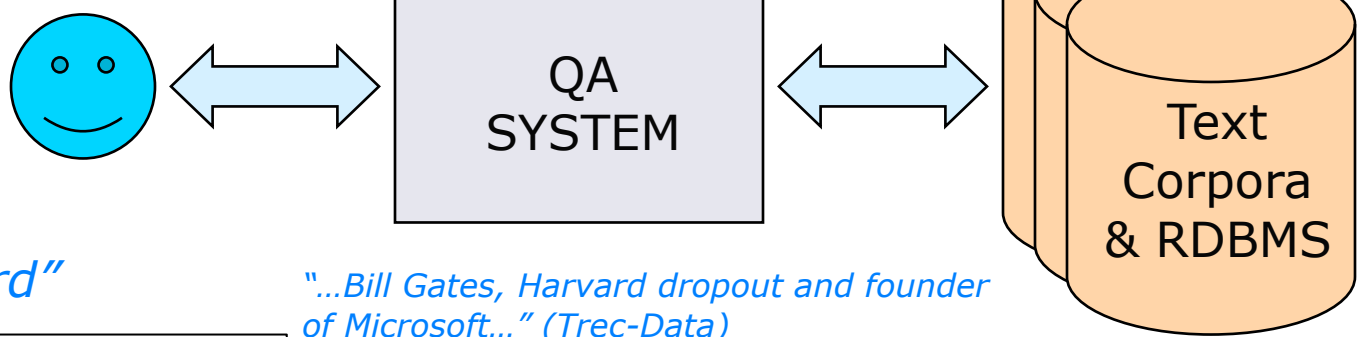
"…Bill Gates, Harvard dropout and founder of Microsoft…" (Trec-Data)

Question Answering

- Input: a question in NL; a set of text and database resources
- Output: a set of possible answers drawn from the resources

"Where did Bill Gates go to college?"

"What is the rainiest place on Earth?"



"Harvard"

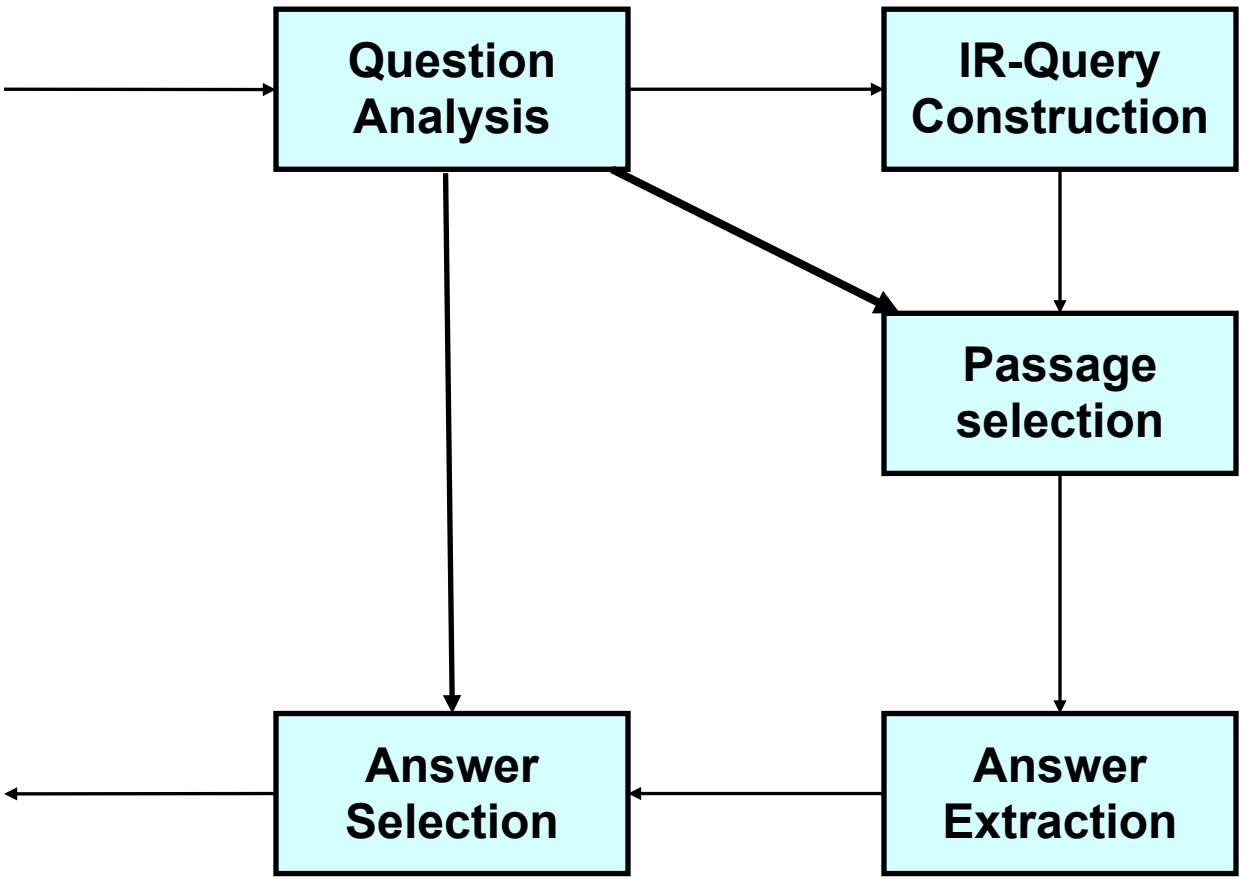
"…Bill Gates, Harvard dropout and founder of Microsoft…" (Trec-Data)

"Mount Waialeale"

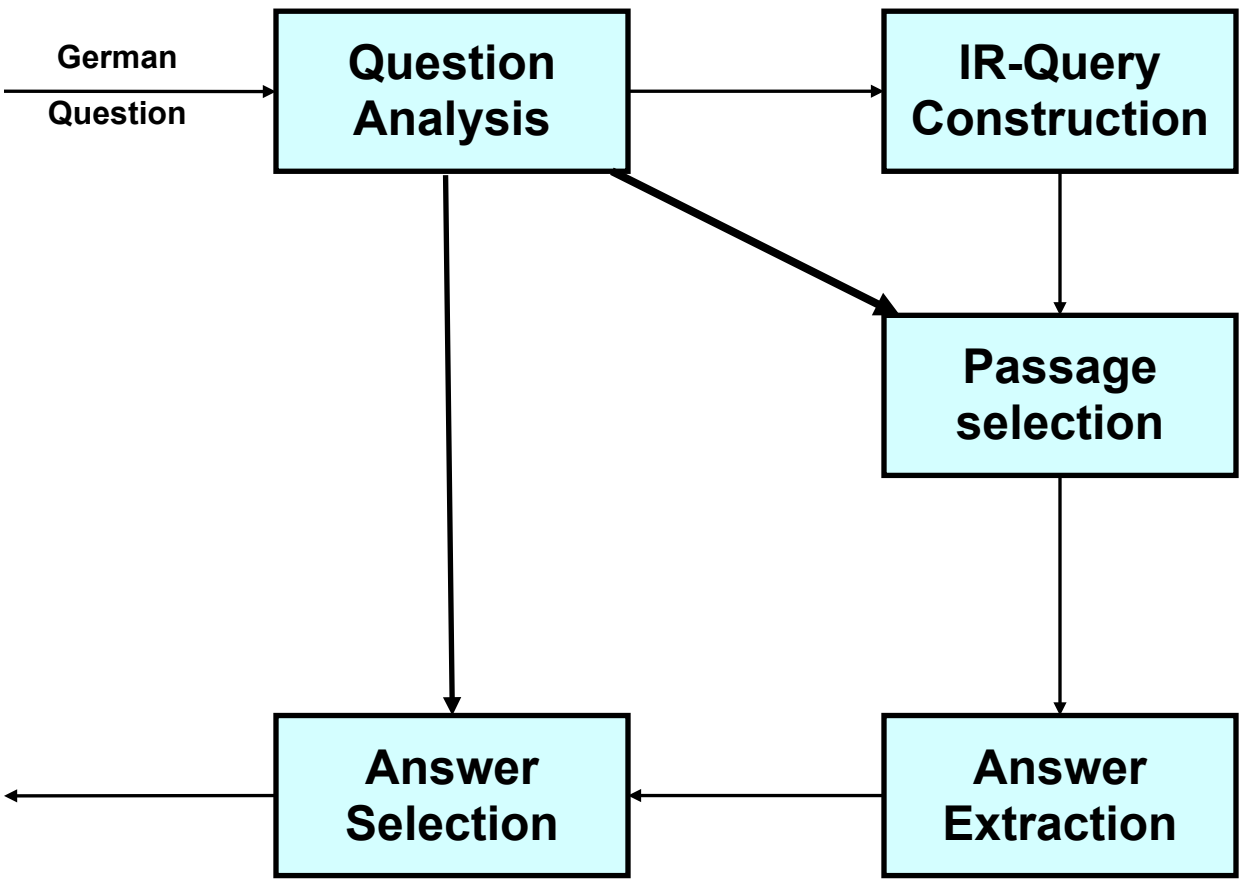
"… In misty Seattle, Wash., last year, 32 inches of rain fell.
Hong Kong gets about 80 inches a year, and even Pago Pago,
noted for its prodigious showers, gets only about 196 inches annually.
(The titleholder, according to the National Geographic Society,
is Mount Waialeale in Hawaii, where about 460 inches of rain falls each year.) …"
(Trec-Data; but see Google-retrieved Web page.)

Open-Domain Question Answering

- Open domain
  - No restriction for the domain and type of question
  - No restriction on document source
- Combines
  - Information retrieval
  - Information extraction
  - Text mining
  - Computational Linguistics
- Cross-lingual ODQA
  - Express query in language X
  - Answer from documents in language Y

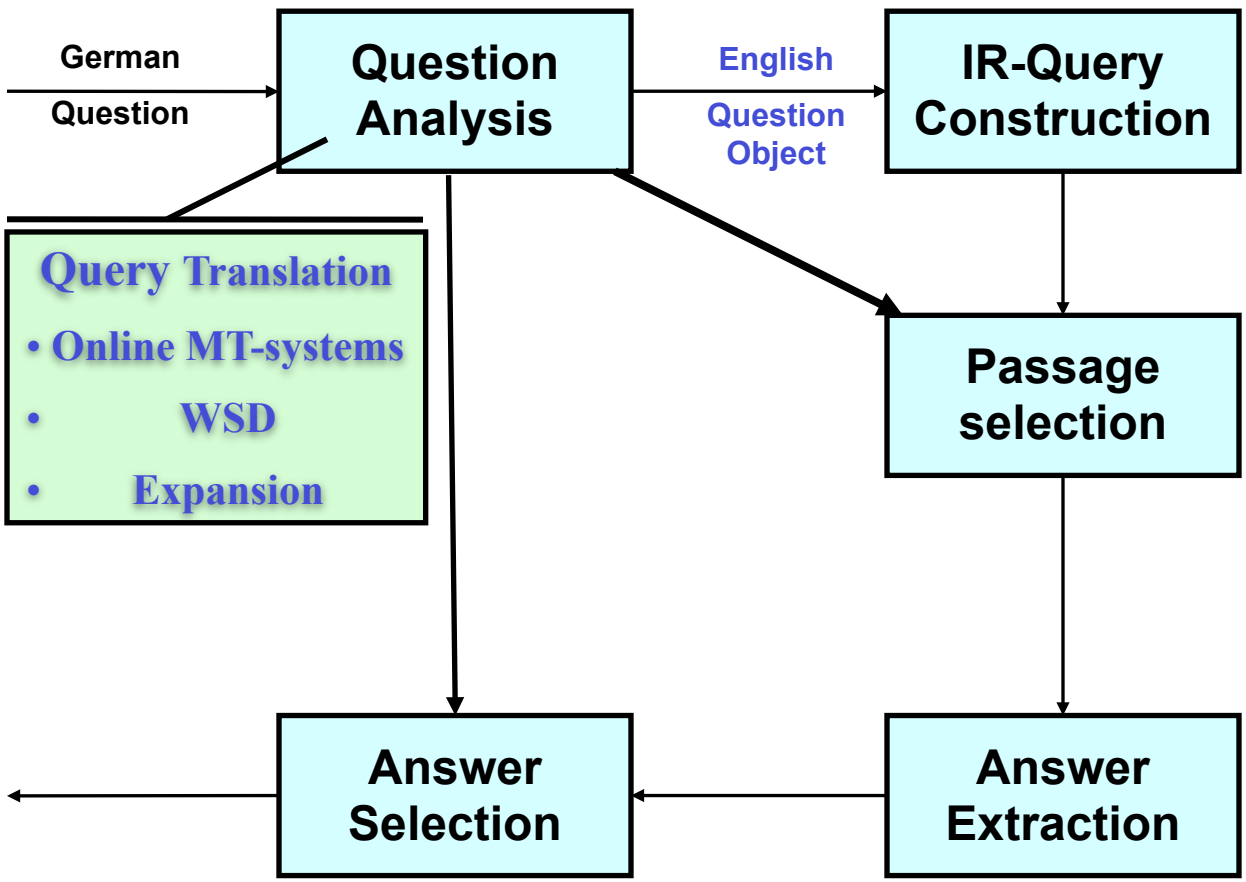

Question Analysis
IR-Query Construction
Passage selection
Answer Selection
Answer Extraction

“Mit wem ist David Beckham verheiratet?”



“Mit wem ist David Beckham verheiratet?”

{person:David Beckham, married, person:?}



"Mit wem ist David Beckham verheiratet?"

{person:David Beckham, married, person:?}

German Question

Question Analysis

English Question Object

IR-Query Construction

IR-Google

IR-Lucene/XML

Annotated Corpus

Documents

Query Translation

- Online MT-systems
- WSD
- Expansion

Passage selection

Answer Selection

Answer Extraction

"Mit wem ist David Beckham verheiratet?"

{person:David Beckham, married, person:?}

German Question

Question Analysis

English Question Object

IR-Query Construction

IR-Google

IR-Lucene/XML

Annotated Corpus

Documents

Query Translation

- Online MT-systems
- WSD
- Expansion

Question Object:

- Focus, Scope
- AnswerType

Passage selection

Passages

"David Beckham, the soccer star engaged to marry Posh Spice, is being blamed for England 's World Cup defeat."

Answer Selection

Answer Extraction

"Mit wem ist David Beckham verheiratet?"

{person:David Beckham, married, person:?}

German Question

Question Analysis

English Question Object

IR-Query Construction

IR-Google

IR-Lucene/XML

Annotated Corpus

Documents

Query Translation
- Online MT-systems
- WSD
- Expansion

Question Object:
- Focus, Scope
- AnswerType

Passage selection

Passages

"David Beckham, the soccer star engaged to marry Posh Spice, is being blamed for England 's World Cup defeat."

Answer Selection

Candidates

Answer Extraction

{person:David Beckham, person:Posh Spice}

"Mit wem ist David Beckham verheiratet?"

{person:David Beckham, married, person:?}

German Question

Question Analysis

English Question Object

IR-Query Construction

IR-Google

IR-Lucene/XML

Annotated Corpus

Documents

Query Translation

- Online MT-systems
- WSD
- Expansion

Question Object:

- Focus, Scope
- AnswerType

Passage selection

Passages

"David Beckham, the soccer star engaged to marry Posh Spice, is being blamed for England 's World Cup defeat."

Answer

Answer Selection

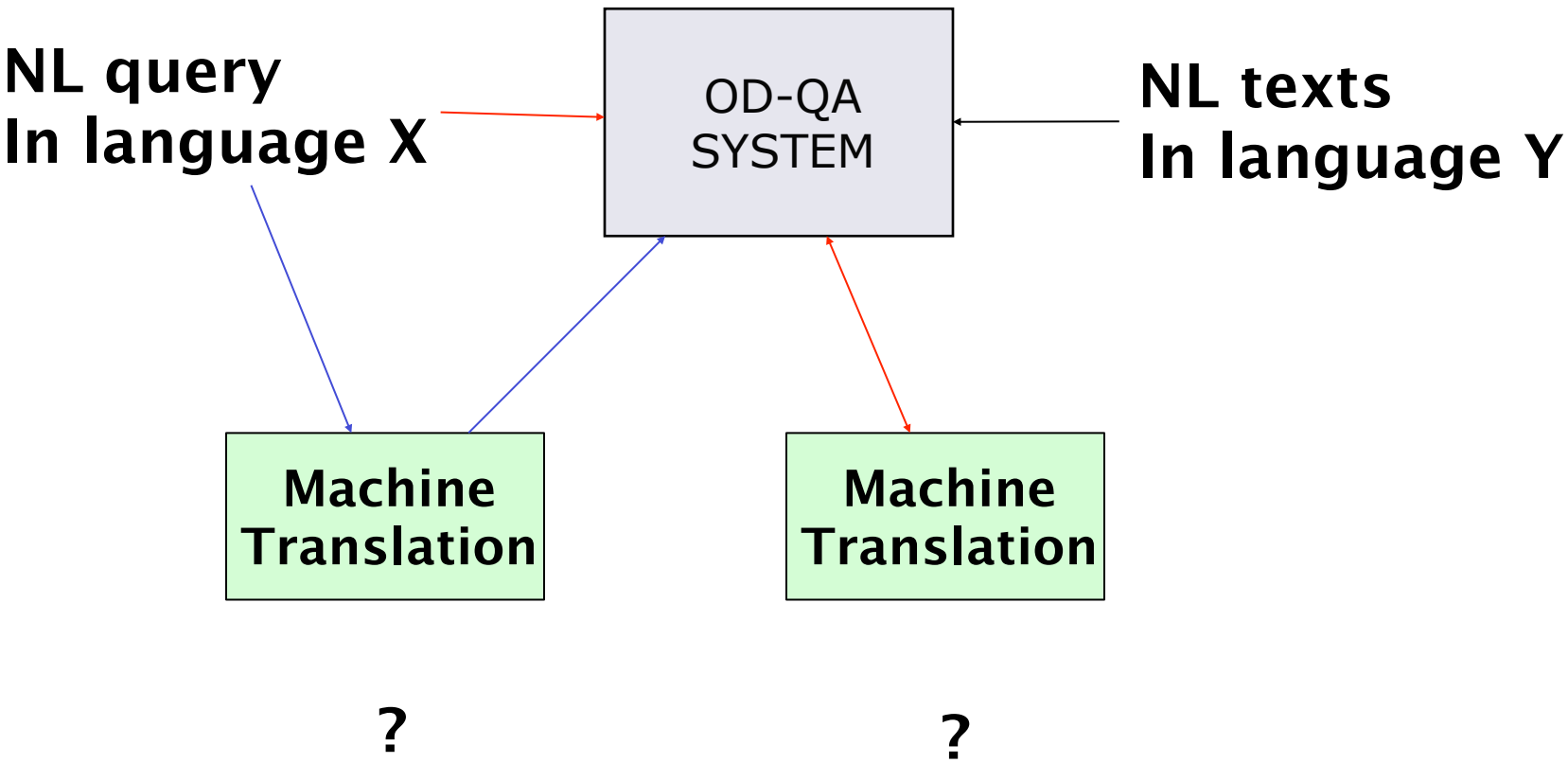
Candidates

Answer Extraction

Posh Spice

{person:David Beckham, person:Posh Spice}

Cross-Lingual ODQA



Approaches in CL QA

Two main different approaches used in Cross-Language QA systems:

1

translation of the question into the target language (i.e. in the language of the document collection)

↓

question processing

↓

answer extraction

2

question processing in the source language to retrieve information (such as keywords, question focus, expected answer type, etc.)

↓

translation and expansion of the retrieved data

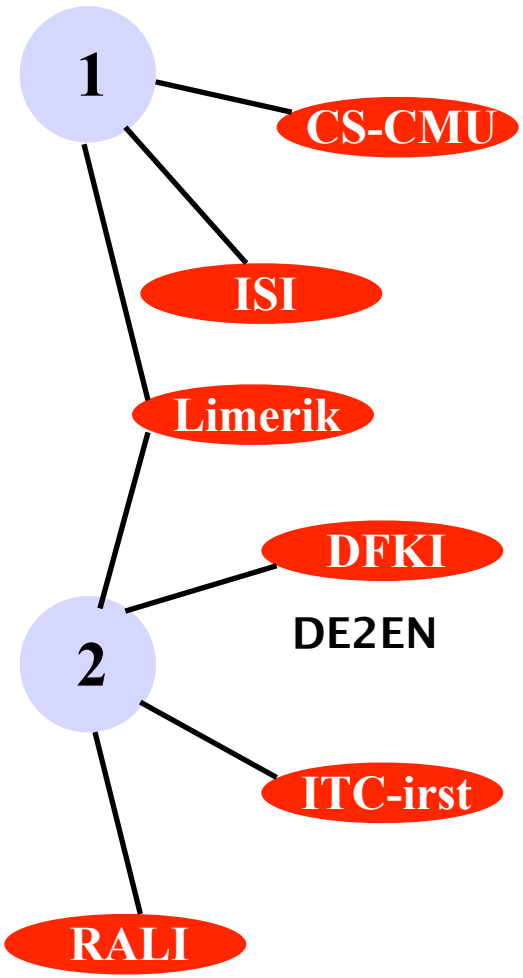
↓

answer extraction

Slide from B. Magnini

Approaches in CL QA

Two main different approaches used in Cross-Language QA systems:



translation of the question into the target language (i.e. in the language of the document collection)

question processing

answer extraction

question processing in the source language to retrieve information (such as keywords, question focus, expected answer type, etc.)

translation and expansion of the retrieved data

answer extraction

Slide from B. Magnini

Query Translation & Expansion

- First idea:
  - Only use EuroWordNet
  - Defines a word-based translation via synset offsets
- Experience
  - EuroWordNet too sparse on German side
  - Neverless introduced too much ambiguity
  - NE-translation is crucial
- So far, not very much of help

Query Translation & Expansion

- First idea:
  - Only use EuroWordNet
  - Defines a word-based translation via synset offsets
- Experience
  - EuroWordNet too sparse on German side
  - Neverless introduced too much ambiguity
  - NE-translation is crucial
- So far, not very much of help

- Second idea:

Query Translation & Expansion

- First idea:
  - Only use EuroWordNet
  - Defines a word-based translation via synset offsets
- Experience
  - EuroWordNet too sparse on German side
  - Neverless introduced too much ambiguity
  - NE-translation is crucial
- So far, not very much of help

- Second idea:
  - Use EuroWordNet

Query Translation & Expansion

- First idea:
  - Only use EuroWordNet
  - Defines a word-based translation via synset offsets
- Experience
  - EuroWordNet too sparse on German side
  - Neverless introduced too much ambiguity
  - NE-translation is crucial
- So far, not very much of help

- Second idea:
  - Use EuroWordNet
  - Use **external** MT-services

Query Translation & Expansion

- First idea:
  - Only use EuroWordNet
  - Defines a word-based translation via synset offsets
- Experience
  - EuroWordNet too sparse on German side
  - Neverless introduced too much ambiguity
  - NE-translation is crucial
- So far, not very much of help

- Second idea:
  - Use EuroWordNet
  - Use **external** MT-services
  - Overlap-mechanism for query expansion

Query Translation & Expansion

- First idea:
  - Only use EuroWordNet
  - Defines a word-based translation via synset offsets
- Experience
  - EuroWordNet too sparse on German side
  - Neverless introduced too much ambiguity
  - NE-translation is crucial
- So far, not very much of help

- Second idea:
  - Use EuroWordNet
  - Use **external** MT-services
  - Overlap-mechanism for query expansion
- Crosslingual because

Query Translation & Expansion

- First idea:
  - Only use EuroWordNet
  - Defines a word-based translation via synset offsets
- Experience
  - EuroWordNet too sparse on German side
  - Neverless introduced too much ambiguity
  - NE-translation is crucial
- So far, not very much of help

- Second idea:
  - Use EuroWordNet
  - Use **external** MT-services
  - Overlap-mechanism for query expansion
- Crosslingual because
  - Q-type & A-type from DE-Question Analysis

Query Translation & Expansion

- First idea:
  - Only use EuroWordNet
  - Defines a word-based translation via synset offsets
- Experience
  - EuroWordNet too sparse on German side
  - Neverless introduced too much ambiguity
  - NE-translation is crucial
- So far, not very much of help

- Second idea:
  - Use EuroWordNet
  - Use **external** MT-services
  - Overlap-mechanism for query expansion
- Crosslingual because
  - Q-type & A-type from DE-Question Analysis
  - Synsets from EuroWN direct query expansion (**online alignment**)

Example

1. Translation services for Word Sense Disambiguation (WSD)

Wo wurde das Militärflugzeug Strike Eagles 1990 eingesetzt?

2. Query expansion using EuroWordNet

Example

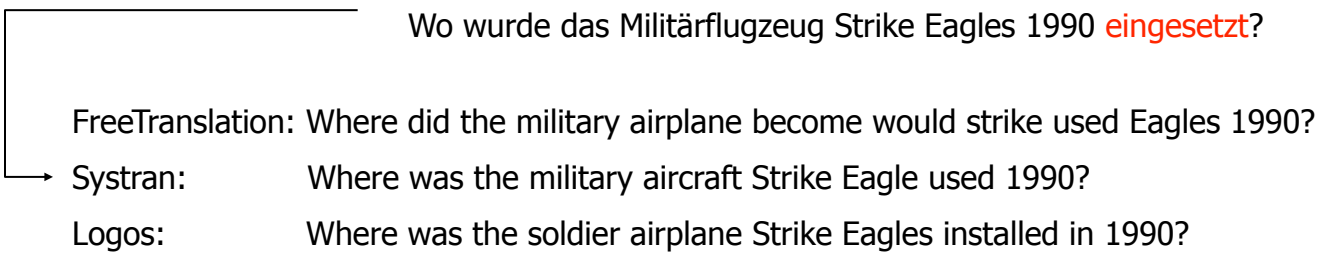
1. Translation services for Word Sense Disambiguation (WSD)

Wo wurde das Militärflugzeug Strike Eagles 1990 eingesetzt?

2. Query expansion using EuroWordNet

Example

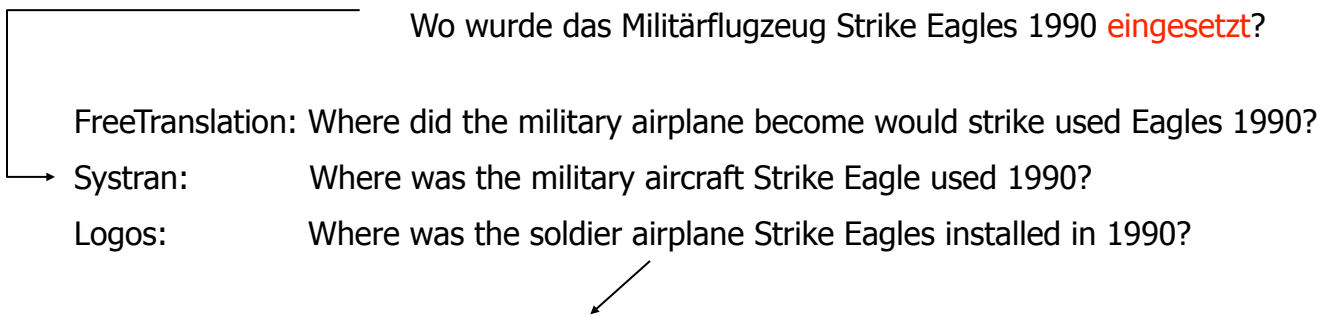
1. Translation services for Word Sense Disambiguation (WSD)



2. Query expansion using EuroWordNet

Example

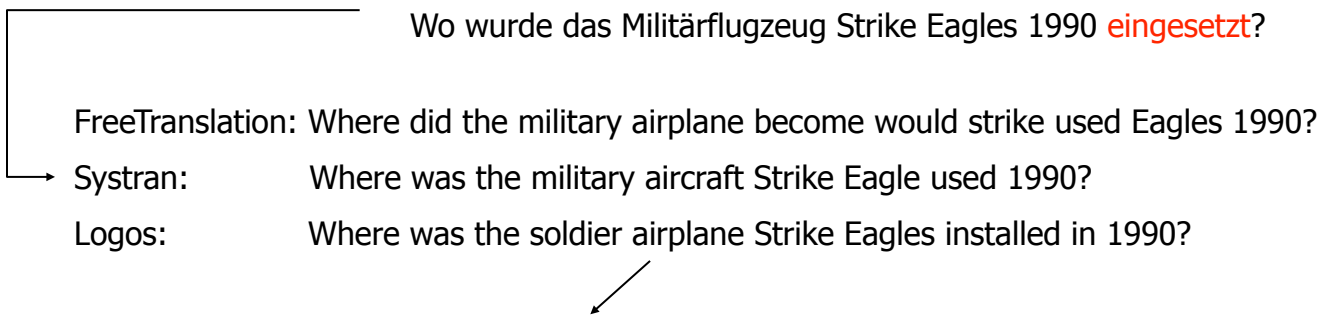
1. Translation services for Word Sense Disambiguation (WSD)



2. Query expansion using EuroWordNet

Example

1. Translation services for Word Sense Disambiguation (WSD)

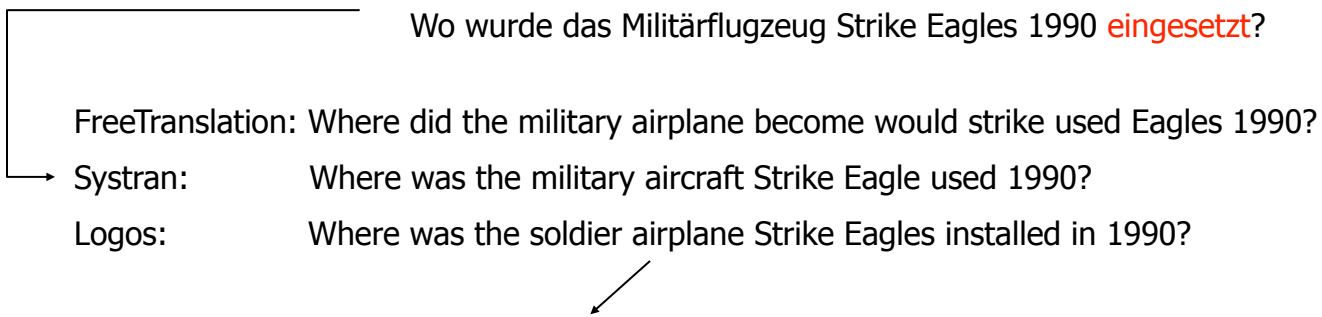


BoO_{EN} := {soldier, airplane, strike, eagle, install, 1990, military, become, strike, use, aircraft}

2. Query expansion using EuroWordNet

Example

1. Translation services for Word Sense Disambiguation (WSD)



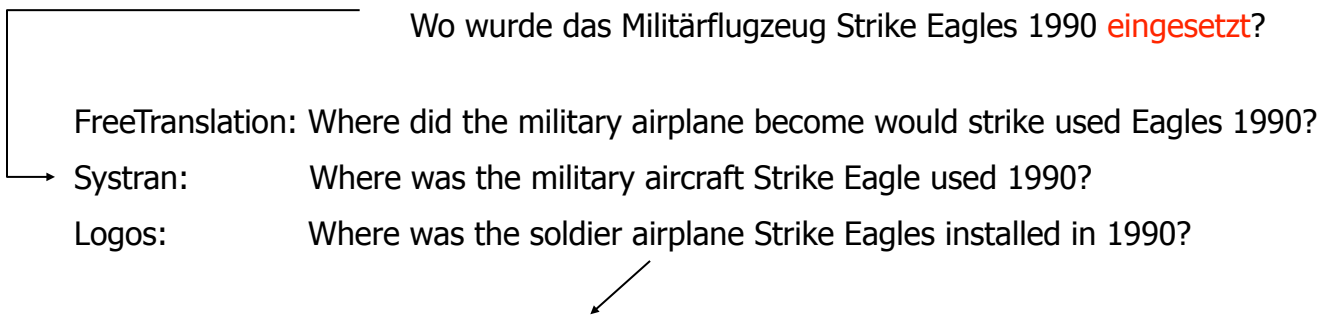
BoO_{EN} := {soldier, airplane, strike, eagle, install, 1990, military, become, strike, use, aircraft}

2. Query expansion using EuroWordNet

∀x ∈ BoO_{EN}: lookup(EuroWN);
If x is unambiguous: extend BoO_{EN}
Else ∀readings(x):
get its aligned German readings &
Look them up in BoO_{GN}
If successfully then add English terms to BoO_{EN}

Example

1. Translation services for Word Sense Disambiguation (WSD)



2. Query expansion using EuroWordNet

$\forall x \in BoO_{EN}$: lookup(EuroWN);
If x is unambiguous: extend BoO_{EN}
Else ∀readings(x):
get its aligned German readings &
Look them up in BoO_{GN}
If successfully then add English terms to BoO_{EN}

Reading-697925

EN: {handle, use, wield}

DE: {handhaben, hantieren}

Example

1. Translation services for Word Sense Disambiguation (WSD)

Wo wurde das Militärflugzeug Strike Eagles 1990 eingesetzt?

FreeTranslation: Where did the military airplane become would strike used Eagles 1990?

Systran: Where was the military aircraft Strike Eagle used 1990?

Logos: Where was the soldier airplane Strike Eagles installed in 1990?

BoO_{EN} := {soldier, airplane, strike, eagle, install, 1990, military, become, strike, use, aircraft}

2. Query expansion using EuroWordNet

$\forall x \in BoO_{EN}$: lookup(EuroWN);
If x is unambiguous: extend BoO_{EN}
Else ∀readings(x):
get its aligned German readings &
Look them up in BoO_{GN}
If successfully then add English terms to BoO_{EN}

Reading-697925
EN: {handle, use, wield}
DE: {handhaben, hantieren}

Reading-1453934:
EN: {behave toward, use}
DE: not aligned

Example

1. Translation services for Word Sense Disambiguation (WSD)

Wo wurde das Militärflugzeug Strike Eagles 1990 eingesetzt?

FreeTranslation: Where did the military airplane become would strike used Eagles 1990?

Systran: Where was the military aircraft Strike Eagle used 1990?

Logos: Where was the soldier airplane Strike Eagles installed in 1990?

BoO_{EN} := {soldier, airplane, strike, eagle, install, 1990, military, become, strike, use, aircraft}

2. Query expansion using EuroWordNet

$\forall x \in BoO_{EN}$: lookup(EuroWN);
If x is unambiguous: extend BoO_{EN}
Else ∀readings(x):
get its aligned German readings &
Look them up in BoO_{GN}
If successfully then add English terms to BoO_{EN}

Reading-697925
EN: {handle, use, wield}
DE: {handhaben, hantieren}

Reading-1453934:
EN: {behave toward, use}
DE: not aligned

Example

1. Translation services for Word Sense Disambiguation (WSD)

Wo wurde das Militärflugzeug Strike Eagles 1990 eingesetzt?

FreeTranslation: Where did the military airplane become would strike used Eagles 1990?

Systran: Where was the military aircraft Strike Eagle used 1990?

Logos: Where was the soldier airplane Strike Eagles installed in 1990?

BoO_{EN} := {soldier, airplane, strike, eagle, install, 1990, military, become, strike, use, aircraft}

2. Query expansion using EuroWordNet

$\forall x \in BoO_{EN}$: lookup(EuroWN);
If x is unambiguous: extend BoO_{EN}
Else ∀readings(x):
get its aligned German readings &
Look them up in BoO_{GN}
If successfully then add English terms to BoO_{EN}

Reading-697925
EN: {handle, use, wield}
DE: {handhaben, hantieren}

Reading-1453934:
EN: {behave toward, use}
DE: not aligned

Reading-658243:
EN: {apply, employ, make use of, put to use, use, utilise, utilize}
DE: {anbringen, anwenden, bedienen, benutzen, einsetzen, ...}

Example

1. Translation services for Word Sense Disambiguation (WSD)

Wo wurde das Militärflugzeug Strike Eagles 1990 eingesetzt?

FreeTranslation: Where did the military airplane become would strike used Eagles 1990?

Systran: Where was the military aircraft Strike Eagle used 1990?

Logos: Where was the soldier airplane Strike Eagles installed in 1990?

BoO_{EN} := {soldier, airplane, strike, eagle, install, 1990, military, become, strike, use, aircraft}

2. Query expansion using EuroWordNet

$\forall x \in BoO_{EN}$: lookup(EuroWN);
If x is unambiguous: extend BoO_{EN}
Else ∀readings(x):
get its aligned German readings &
Look them up in BoO_{GN}
If successfully then add English terms to BoO_{EN}

Reading-697925
EN: {handle, use, wield}
DE: {handhaben, hantieren}

Reading-1453934:
EN: {behave toward, use}
DE: not aligned

Reading-658243:
EN: {apply, employ, make use of, put to use, use, utilise, utilize}
DE: {anbringen, anwenden, bedienen, benutzen, einsetzen, ...}

Find documents written in any language

– Using queries expressed in a single language

Source: D. Oard Cross-Language IR presentation

CLIR Conferences

- ☆ Cross Language Evaluation Forum (CLEF)
  - CLIR using European languages.
    - Bulgarian, Danish, Dutch, English, Finnish, French, German, Italian, Portuguese, Spanish, Swedish, Russian
    - http://clef.iei.pi.cnr.it/

- ☆ NTCIR (NII-NACSIS Test Collection for IR Systems) Project
  - CLIR in Asian Languages
    - Chinese, Japanese, and Korean
    - http://research.nii.ac.jp/ntcir/index-en.html

Multilingual QA Track at Clef

	2003	2004	2005	2006	2007/2008
Target languages	3	7	8	9	10
Collections	News 1994		+News 1995		+Wikipedia Nov. 2006
Type of questions	200 Factoid		+ temporal restrictions + Definitions	-Type of questions + Lists	+ Linked questions + Closed lists
Supporting information	Doc.	Doc.	Doc.	Snippet	Snippet
Pilots and exercises		-Temporal restrictions - Lists		-AVE - RealTime - WiQA	- AVE - QAST

Clef 2006: 200 Questions

- FACTOID (150): loc, mea, org, oth, per, tim
- DEFINITION (40): per, org, object, oth
  - Person: Who is Josef Paul Kleihues?
  - Object: What is a router?
  - Other: What is a tsunami?
- LIST (10): “Name works by Tolstoy.”
- Temporally restricted (40): by date, by period, by event
- NIL questions (without known answer in the collection
- Input format: question type (F, D, L) not indicated

Clef 2007: Clef 2006 plus

- ☆ Closed lists:
  - Who were the components of the Beatles?
  - Who were the **last three** presidents of Italy?
- ☆ Linked questions
  - Topic: Otto von Bismarck
    - Who was called the “Iron-Chancellor”?
    - When was he born?
    - Who was his first wife?
  - Topics
    - Person or Event
    - Not provided to participants
    - Only a portion of the questions (from 15% depending on the languages)

Run format

- Clef 2006:
  - Multiple answers: from one to ten *exact* answers per question
  - *exact* = neither more nor less than the information required
  - each answer has to be supported by
    - docid
    - one to ten text snippets justifying the answer (substrings of the specified document giving the actual context)
- Clef 2007:
  - News articles
  - Wikipedia dump from November 2006 (→ caused critical decrase of performance)

Clef 2007: Activated Tasks

(at least one registered participant)

- 10 Source languages (11 in 2006, 10 in 2005)
- 9 Target languages (8 in 2006, 9 in 2005)

S / T	BG	DE	EN	ES	FR	IN	IT	NL	PT	RO
BG										
DE										
EN										
ES										
FR										
IT										
NL										
PT										
RO										

Activated Tasks

- questions were not translated in all the languages
- **Gold Standard**: questions in multiple languages only for tasks were there was at least one registered participant

	MONOLINGUAL	CROSS-LINGUAL	TOTAL
CLEF 2003	3	5	8
CLEF 2004	6	13	19
CLEF 2005	8	15	23
CLEF 2006	7	17	24
CLEF 2007	8	29	37

LT-Lab

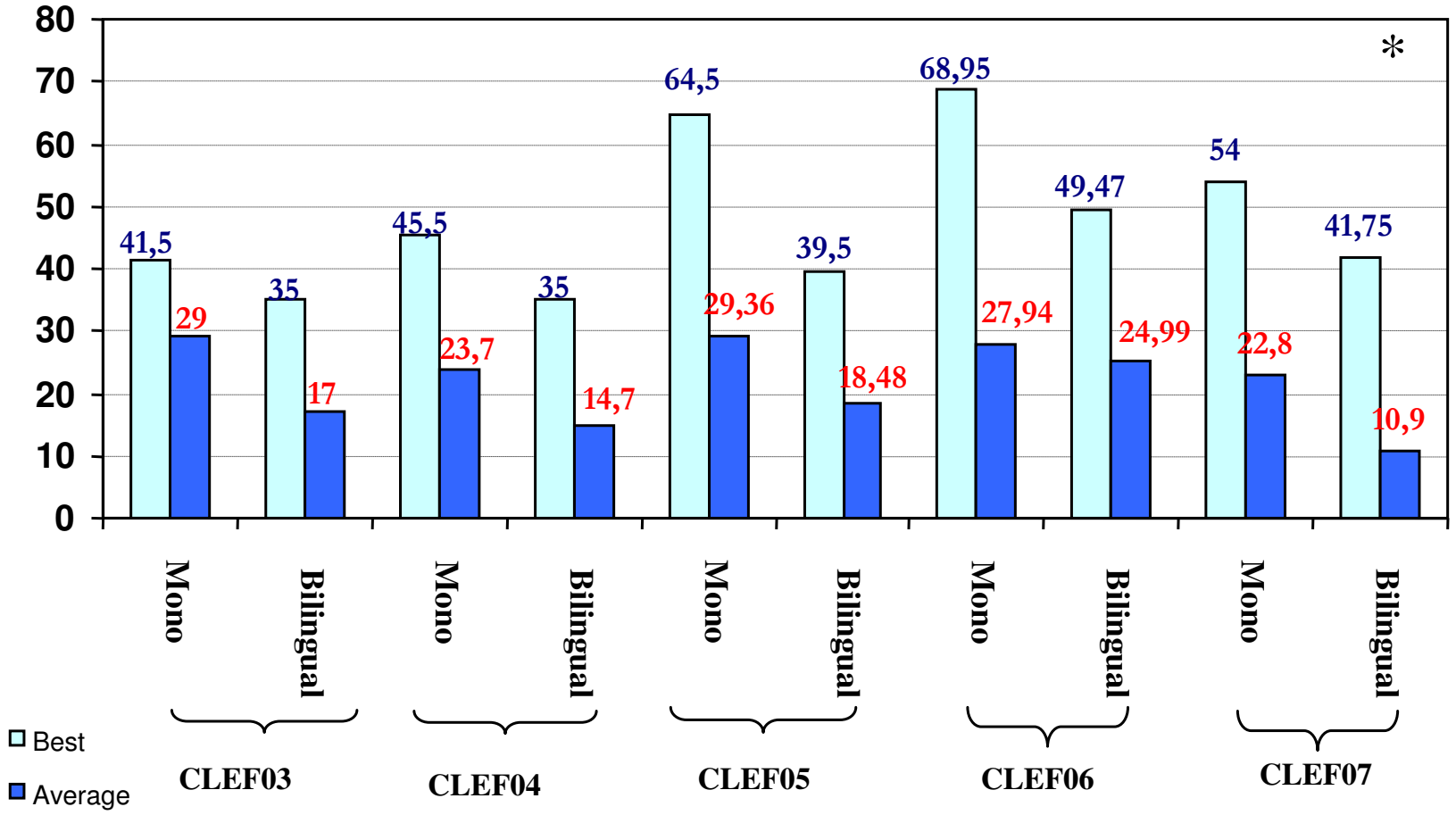
List of participants (2006, 2007)

Acronym	NAME	Contry
SYNAPSE	SYNAPSE Developpement	France
Ling-Comp	U.Rome-La Sapienza	Italy
Alicante	U.Alicante- Informatica	Spain
Hagen	U.Hagen-Informatics	Germany
Daedalus	Daedalus Consortium	Spain
Jaen	U.Jaen-Intell.Systems	Spain
ISLA	U.Amsterdam	Netherlands
INAOE	Inst.Astrophysics,Optics&Electronics	Mexico
DEPOK	U.Indonesia-Comp.Sci.	Indonesia
DFKI	DFKI-Lang.Tech.	Germany
FURUI Lab.	Tokyo Inst Technology	Japan
Linguateca	Linguateca-Sintef	Norway
LIC2M-CEA	Centre CEA Saclay	France
LINA	U.Nantes-LINA	France
Priberam	Priberam Informatica	Portugal
U.Porto	U.Porto- AI	Portugal
U.Groningen	U.Groningen-Letters	Netherlands
	Univ. of Evora	Portugal
	Univ. Poli. De Catalunay	Spain

Acronym	NAME	Country
Lab.Inf.D' Avignon	Lab.Inf. D'Avignon	France
U.Sao Paulo	U.Sao Paulo – Math	Brazil
Vanguard	Vanguard Engineering	Mexico
LCC	Language Comp. Corp.	USA
UAIC	U.AI.I Cuza" Iasi	Romania
Wroclaw U.	Wroclaw U.of Tech	Poland
RFIA-UPV	Univ.Politècnica de Valencia	Spain
LIMSI	CNRS Lab-Orsay Cedex	France
U.Stuttgart	U.Stuttgart-NLP	Germany
FBK	FBK-IRST	Italy
JRC-ISPRA	Institute for the Protection and the Security of the Citizen	Italy
BTB	BulTreeBank Project	Bulgaria
dltg	University of Limerick	Ireland
	INESC-ID	Portugal
	Univ. Wolverhampton	UK
	Cindi Group	Canada
	Macquarie University	Australia
	RACAI	Romania

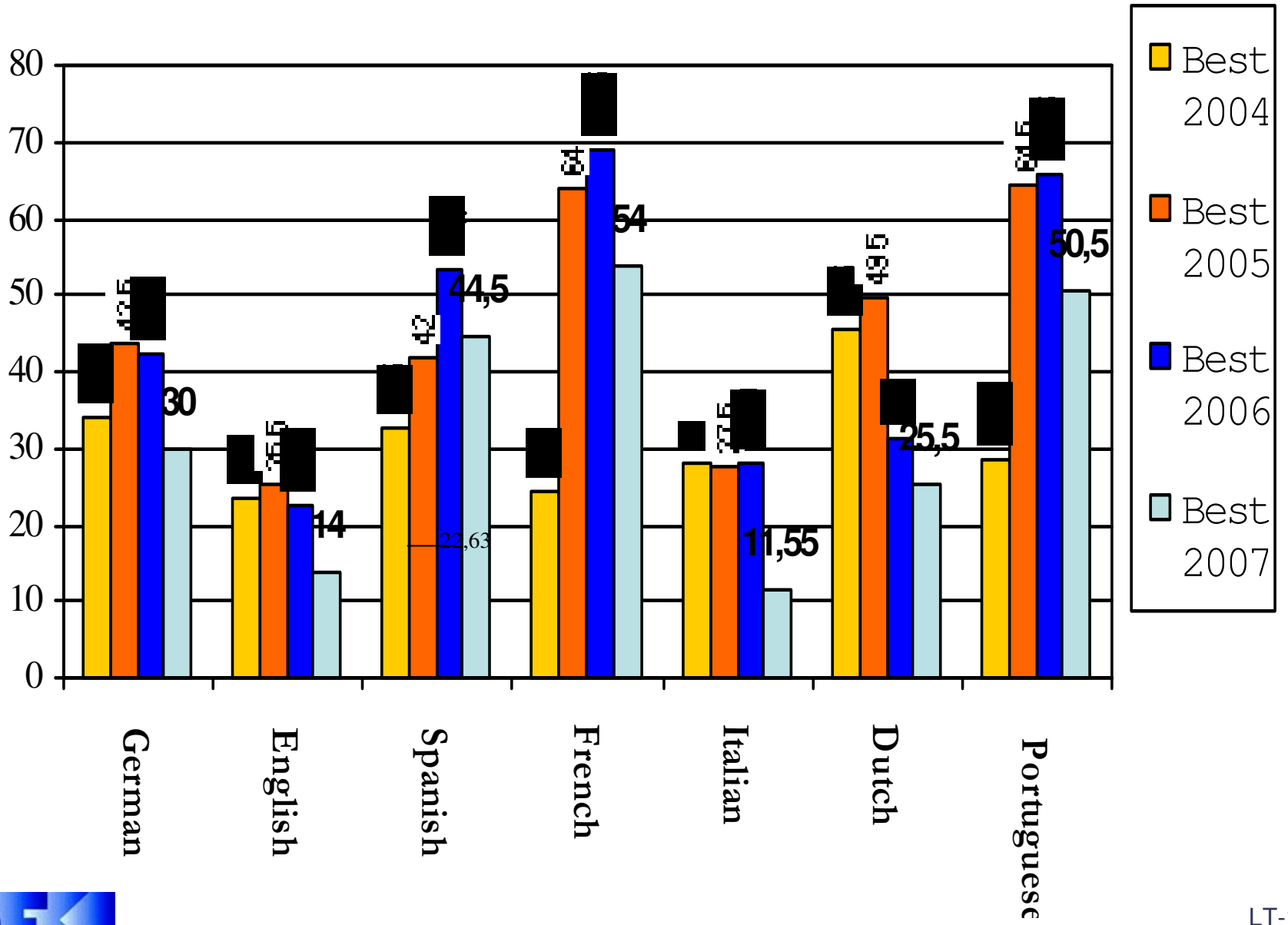
BLUE=Industrial Companies, GREEN=2006 + 2007, RED=not 2007, BLACK=new 2007

Results: Best and Average scores



* This result is still under validation.

Best results in 2004, 2005, 2006, 2007



LT-Lab Participants in 2004 - 2007: compared best results