

Speech and Music Retrieval

LBSC 796/INFM 718R

Session 11, April 20, 2011

Douglas W. Oard

Audio Indexing

- Music
- Sounds
- Speech

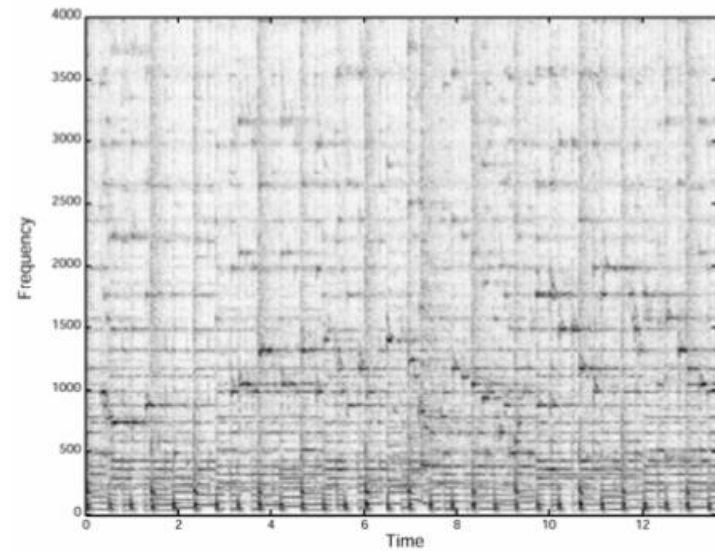
Encoding Audio

- Nyquist theorem: sampling rate
- Channel characteristics
 - Bandpass
 - Amplitude resolution
- Compression
 - Lossless
 - Perceptually-based lossy (e.g., .mp3)

User-Centered Music Retrieval

- User types:
 - Casual
 - name that tune, playlist generation, query by humming, ...
 - Professional
 - Instrument detection, timbre, ...
 - Scholar
- Content types:
 - Score
 - Performance

Shazam



- Spectrogram
 - Time vs. frequency plot of intensity
- Peak filter
 - Local maxima of intensity
- Difference coding
 - Rate of frequency change over time
- Convolution
 - Sliding correlation

Muscle Fish Audio Retrieval

- Compute 4 acoustic features for each time slice
 - Pitch, amplitude, brightness, bandwidth
- Segment at major discontinuities
 - Find average, variance, and smoothness of segments
- Store pointers to segments in 13 sorted lists
 - Use a commercial database for proximity matching
 - 4 features, 3 parameters for each, plus duration
 - Then rank order using statistical classification
- Display file name and audio

New Zealand Melody Index

- Index musical tunes as contour patterns
 - Rising, descending, and repeated pitch
 - Note duration as a measure of rhythm
- Users sing queries using words or la, da, ...
 - Pitch tracking accommodates off-key queries
- Rank order using approximate string match
 - Insert, delete, substitute, consolidate, fragment
- Display title, sheet music, and audio

Contour Matching Example

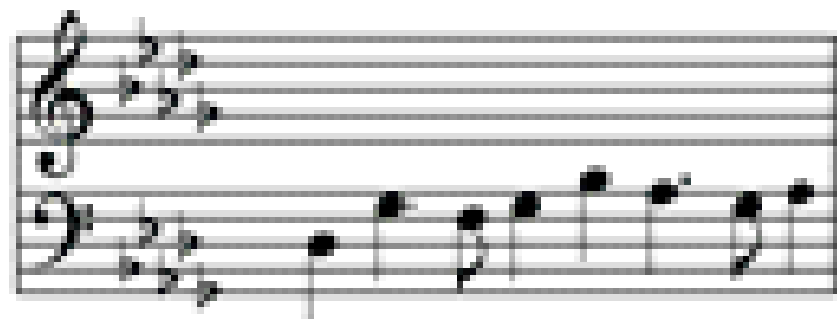
- “Three Blind Mice” is indexed as:
 - *DDUDDUDRDUDRD
 - * represents the first note
 - D represents a descending pitch (U is ascending)
 - R represents a repetition (detectable split, same pitch)
- My singing produces:
 - *DDUDDUDRRUDRR
- Approximate string match finds 2 substitutions



What you
sang



What I
heard



QUERY RESULTS

Your melody matched 609 tunes

[Search again](#)

[Choose sound format](#)

| Tune | | Database | Match |
|------|---|----------------|-------|
| | ACH HAMELN DU WUNDERSCHÖNE STADT | German | 100% |
| | Auld Lang Syne | North American | 100% |
| | Beli derhi shi qionghen | Chinese | 100% |
| | BLUEHE LIEBES VEILCHEN DAS ICH SEBST ERZOG | German | 100% |
| | CHRISTINCHEN OHEM IM GARTEN | German | 100% |
| | Das böse Stiefmütterlein - Es ging ein treuerliebtes Paar | German | 100% |
| | Das böse Stiefmütterlein - Es war einmal ein Kindleinlein Mein, | German | 100% |



Muscle Fish LLC

**Specializing in Audio,
Music, MIDI, DSP and
Audio Search
Technologies.**

[Current projects](#)

[Products](#)

[0.00] laughter/laughterYoungMale
[0.23] laughter/laughterYoungMale3
[0.24] laughter/laughterYoungMale4
[0.35] laughter/laughterFemale1
[0.41] laughter/laughterFemale3
[0.45] laughter/laughterYoungMale2
[0.49] laughter/laughterFemale2
[0.50] mcgill/percussion/sbonogroll
[0.57] animals/dogsBark2
[0.60] telephone/telephoneTouchTones1
[0.61] machines/hammering2
[0.62] mcgill/percussion/lbongoroll
[0.62] speech/female/female3.professor
[0.62] animals/chickenCoop2
[0.62] telephone/telephoneTouchTones2

Sounds in order of similarity

Play the selected sounds

Reorder the sounds by
similarity to the selected sounds

Clear your selections

Spoken Word Collections

- Broadcast programming
 - News, interview, talk radio, sports, entertainment
- Scripted stories
 - Books on tape, poetry reading, theater
- Spontaneous storytelling
 - Oral history, folklore
- Incidental recording
 - Speeches, oral arguments, meetings, phone calls

Some Statistics

~250,000 hours of oral history in British Library

>4,000 radio stations webcasting

~35 million hours per year

~100 billion hours of phone calls each year

Economics of the Web

| | Text in 1995 | Speech in 2003 |
|--|----------------|--------------------------|
| Storage (words per \$) | 300K | 1.5M |
| Internet Backbone (simultaneous users) | 250K | 30M |
| Modem Capacity (% utilization) | 100% | 20% |
| Display Capability (% US population) | 10% | 38% |
| Search Systems | Lycos Yahoo | SpeechBot SingingFish |

Audio Results

1 - 10 of about 135 for **nasa apollo** - 0.38 sec.Search: [Online Music Stores](#) | [the Web](#) | **Podcasts**[Wide Awake in the Sea of Tranquillity](#)http://science.nasa.gov/headlines/y2006/19jul_seaoftranquillity.htm - last published: Wed, 19 Jul 2006 00:00:00 CST  

The fourth installment of Science@NASA's **Apollo** Chronicles explains why Neil Armstrong and Buzz Aldrin couldn't fall asleep in the Sea of Tranquillity. Please vote for this podcast at PodcastAlley! Get this podcast story.

[Wide Awake in the Sea of Tranquillity](#)http://science.nasa.gov/headlines/y2006/19jul_seaoftranquillity.htm - last published: Wed, 19 Jul 2006 00:00:00 CST  

The fourth installment of Science@NASA's **Apollo** Chronicles explains why Neil Armstrong and Buzz Aldrin couldn't fall asleep in the Sea of Tranquillity. Please vote for this podcast at PodcastAlley! Get this podcast story.

[vSocial Video - Apollo 15 Proves Galileo Right](#)<http://www.vsocial.com/video/?d=40719> - last published: Tue, 01 Aug 2006 03:55:35 -0600  

...that the homeward journey was based critically on the validity of the particular theory being tested. Joe Allen, **NASA** SP-289, **Apollo 15** Preliminary Science Report, Summary of Scientific Results, p. 2-11

[RSS - NASA Truth Embargo](#)<http://www.richardsyrett.com> - last published: Wed, 16 Aug 2006 08:59:48 -0500  

Victor Viggiani from Exopolitics Toronto discusses Astronaut Buzz Aldrin coming forward to challenge **NASA** secrecy surrounding **Apollo 11's** sighting of a UFO.

[Episode 19 - Mysterious Universe](#)<http://feeds.feedburner.com/mysteriousuniverse?m=16> - last published: Sun, 29 Jan 2006 23:08:21 +1100  

...listeners. Mexico - Controversy Rages Over Yucatan 'Alien' Fox TV and the **Apollo** Moon Hoax The Daily Planet Debunks the Moon Hoax Theory (Video) Climate Expert Says **NASA** Tried to Silence Him Russia to Open Moonbase Mine Variant Frequencies...

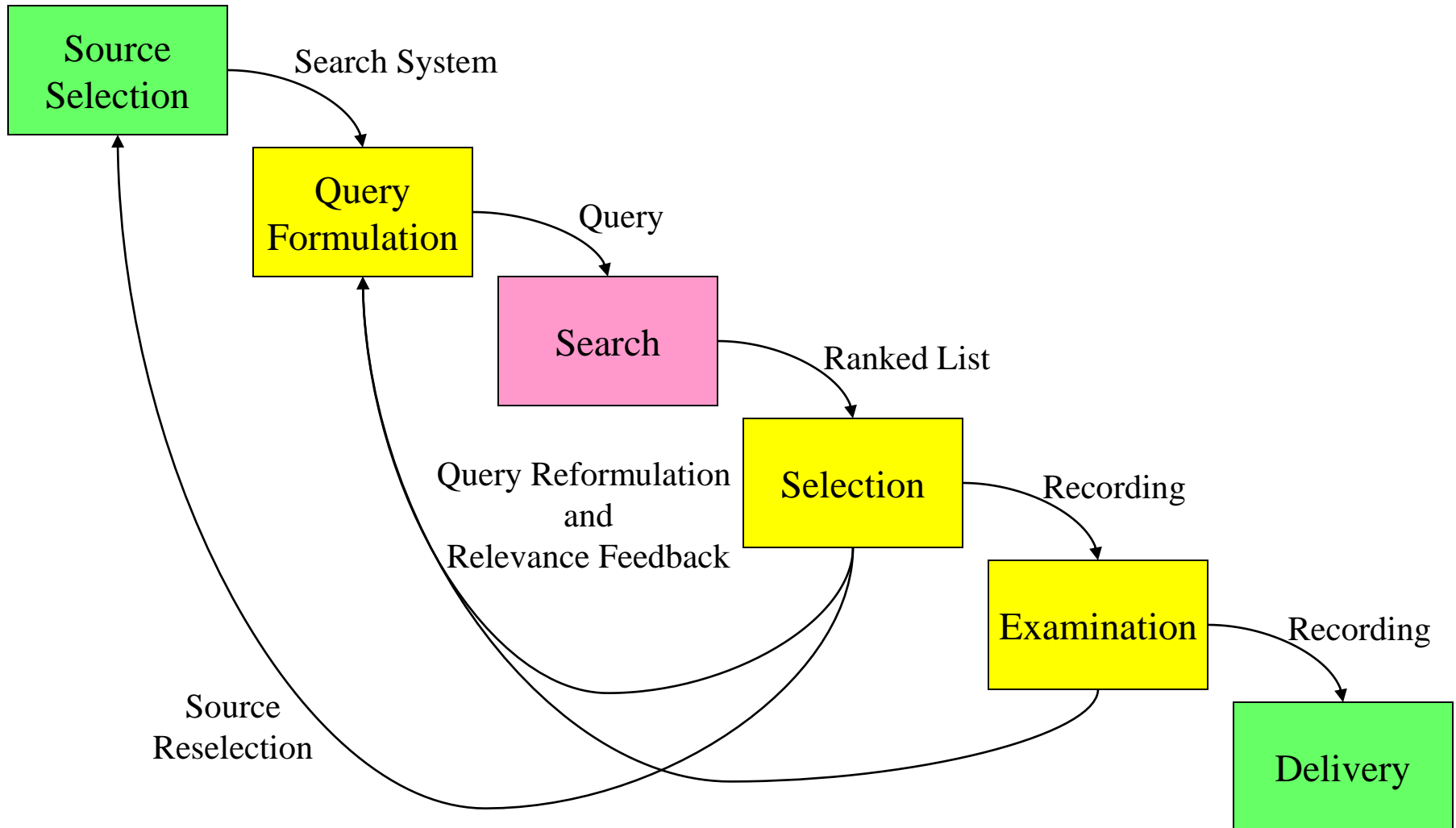
Speech Retrieval Approaches

- Controlled vocabulary indexing
- Ranked retrieval based on associated text

⇒ **Automatic feature-based indexing**

- Social filtering based on other users' ratings

Supporting Information Access



Description Strategies

- Transcription
 - Manual transcription (with optional post-editing)
- Annotation
 - Manually assign descriptors to points in a recording
 - Recommender systems (ratings, link analysis, ...)
- Associated materials
 - Interviewer's notes, speech scripts, producer's logs
- Automatic
 - Create access points with automatic speech processing

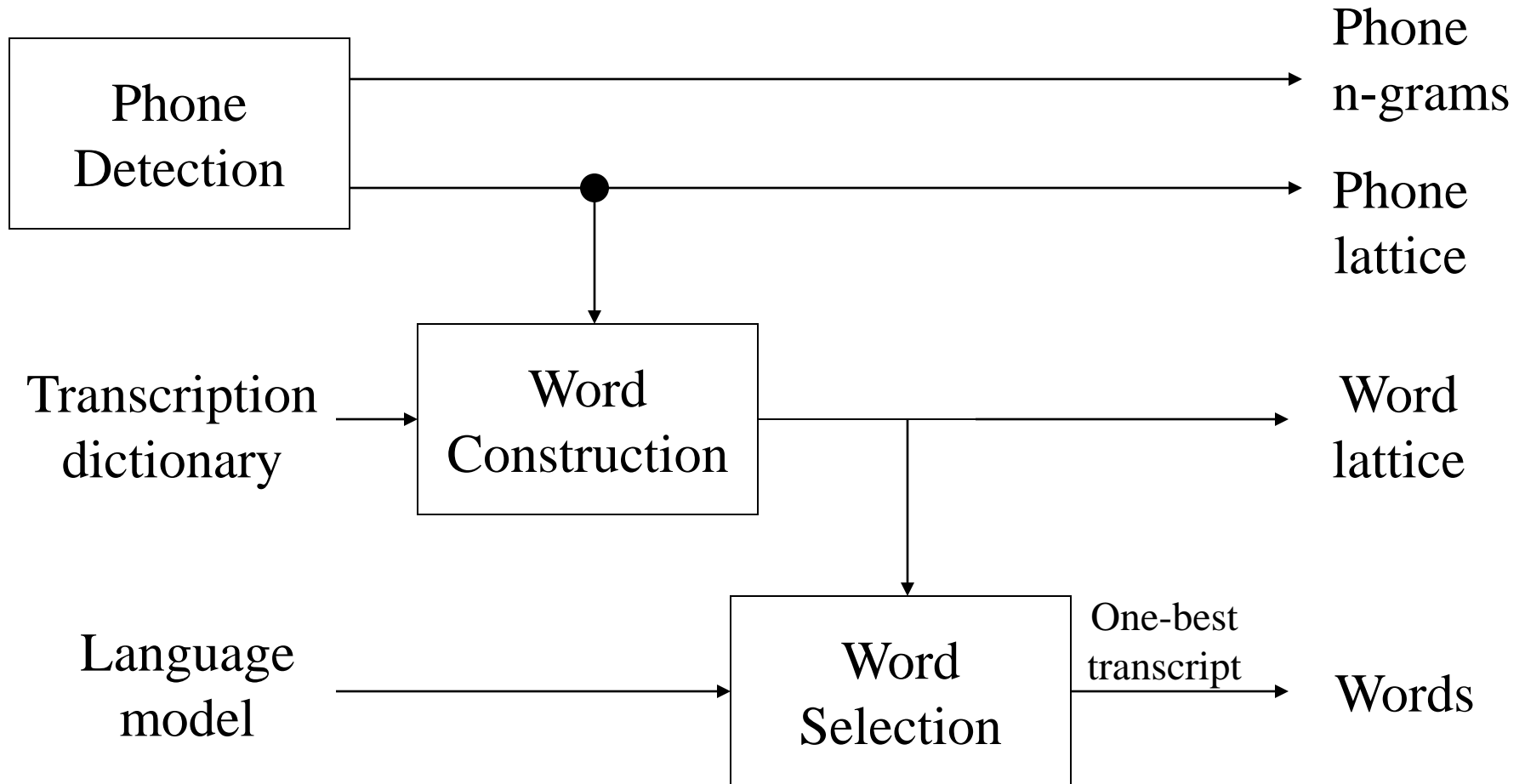
Detectable Speech Features

- Content
 - Phonemes, one-best word recognition, n-best
- Identity
 - Speaker identification, speaker segmentation
- Language
 - Language, dialect, accent
- Other measurable parameters
 - Time, duration, channel, environment

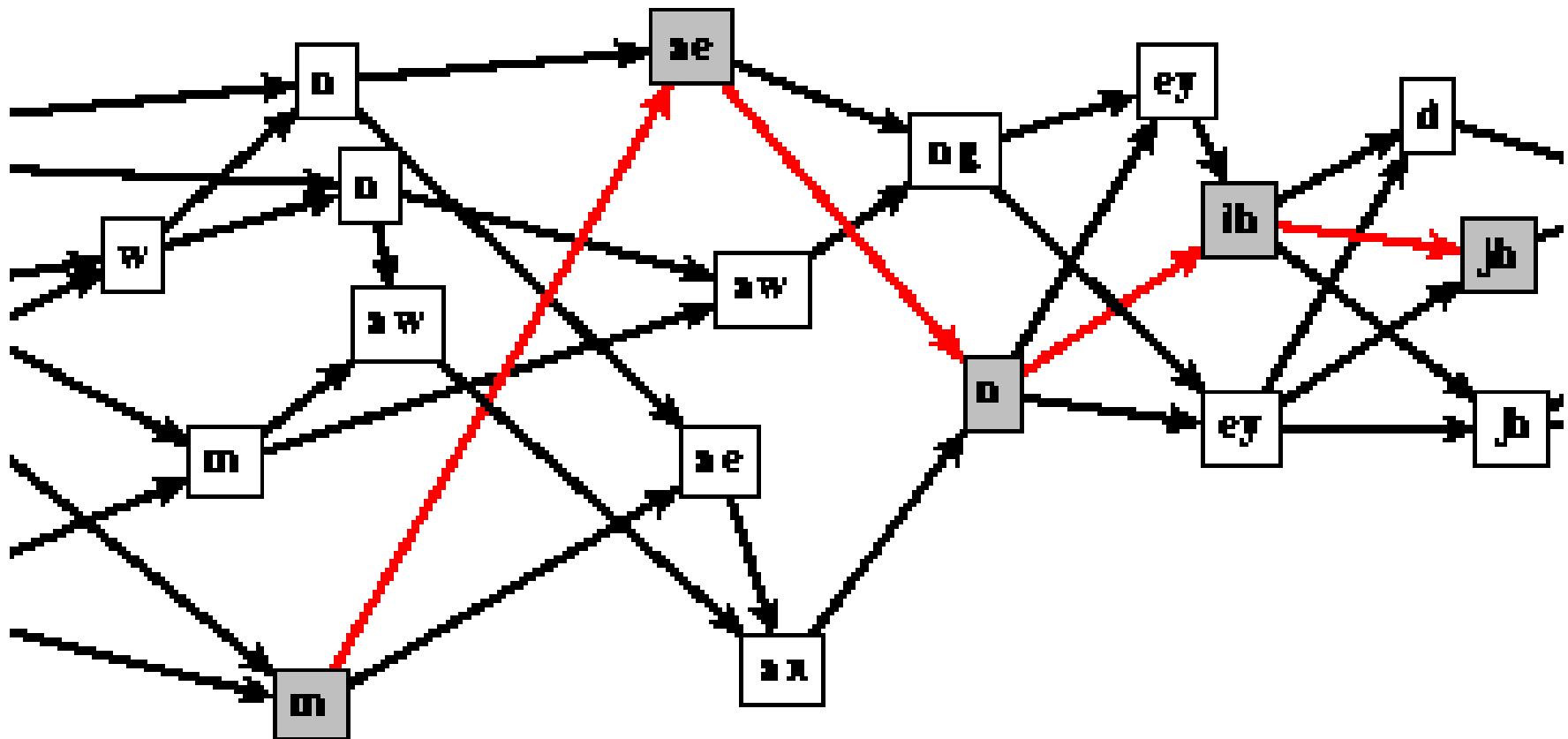
How Speech Recognition Works

- Three stages
 - What sounds were made?
 - Convert from waveform to subword units (phonemes)
 - How could the sounds be grouped into words?
 - Identify the most probable word segmentation points
 - Which of the possible words were spoken?
 - Based on likelihood of possible multiword sequences
- These can be learned from “training data”
 - Hill-climbing sequence models
 - (a “Hidden Markov Model”)

Using Speech Recognition



Phone Lattice



Phoneme Trigrams

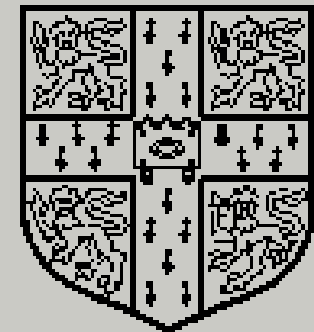
- Manage -> m ae n ih jh
 - Dictionaries provide accurate transcriptions
 - But valid only for a single accent and dialect
 - Rule-base transcription handles unknown words
- Index every overlapping 3-phoneme sequence
 - m ae n
 - ae n ih
 - n ih jh

Cambridge Video Mail Retrieval

- Translate queries to phonemes with dictionary
 - Skip stopwords and words with ≤ 3 phonemes
- Find no-overlap matches in the lattice
 - Queries take about 4 seconds per hour of material
- Vector space exact word match
 - No morphological variations checked
 - Normalize using most probable phoneme sequence
- Select from a ranked list of subject lines

Video Mail Messages:

| | | | | |
|-----|-------|----------|---------------------|------------------------|
| mgb | 10:00 | 96/05/14 | beer festival as we | <div><div></div></div> |
| fb | 10:17 | 96/05/16 | folk fest folks | <div><div></div></div> |
| mgb | 17:34 | 96/05/10 | Stop button and Cam | <div><div></div></div> |
| mda | 17:15 | 96/05/10 | (VMR) Active Clothi | <div><div></div></div> |
| mgb | 15:52 | 96/05/13 | Parkside redevelopm | <div><div></div></div> |
| mda | 16:50 | 96/05/10 | (VMR) A question on | <div><div></div></div> |
| mda | 16:37 | 96/05/10 | (VMR) Digital (TI-m | <div><div></div></div> |
| mda | 16:18 | 96/05/07 | (VMR) MME167 syste | <div><div></div></div> |
| fb | 10:20 | 96/05/16 | make money fast! | <div><div></div></div> |
| mda | 17:04 | 96/05/10 | (VMR) Diary and Sto | <div><div></div></div> |
| fms | 19:27 | 96/05/09 | plugins, monolithic | <div><div></div></div> |



Display by:

☒ score☐ user☐ date

Submit

Request: folk festival cambridge

Clear

Record

Play

Browse

Delete

Quit

File Edit Debug

olivetti



☒ On/Off

☐ Mute

Audio Volume

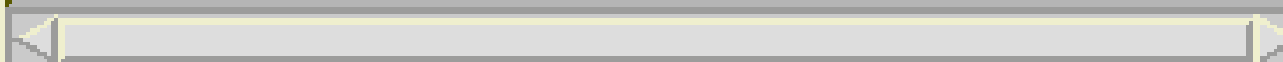


10

festival



10



<<

<

>

>>

Playback

Pause

Stop

Cancel

Key Results from TREC/TDT

- Recognition and retrieval can be decomposed
 - Word recognition/retrieval works well in English
- Retrieval is robust with recognition errors
 - Up to 40% word error rate is tolerable
- Retrieval is robust with segmentation errors
 - Vocabulary shift/pauses provide strong cues

Competing Demands on the Interface

- Query must result in a manageable set
 - But users prefer simple query interfaces
- Selection interface must show several segments
 - Representations must be compact, but informative
- Rapid examination should be possible
 - But complete access to the recordings is desirable

BBN Radio News Retrieval

Your query has finished

Search

Clear

☐ OR
☒ AND

Topic

Organization








Speaker

Story



Person

Location



Text





Rough'n'Ready **GTE**



5 stories about: Jewish-Arab relations : Politics and government : Palestinian Arabs : Middle East : Israel : Terr



Jewish-Arab relations : Politics and government : Palestinian Arabs : Middle East : Israel : Terrorism : Pale



Jewish-Arab relations : Israel : Middle East : Middle East peace negotiations : Politics and government : P

male 5

Well as all work during president Clinton's trip to New York tonight and he enjoys the performance of the opera Carmen at Lincoln Center and see the scene there is a lot of Broadway. Now earlier today Mr. Clinton announced that the UN united Nations general assembly that he plans to send a nuclear test ban treaty to the Senate the treaty bans all nuclear test explosions and is regarded as a milestone in the arms control. Two israeli security guards were wounded in an early morning shooting in Jordan a government official says three men and a car opened fire on the guard's car wounding both before Skipping guards were treated at a hospital and released it is real several West Bank villages were sealed by israeli soldiers who search for the islamic militants behind two recent suicide bombings in Jerusalem palestinian leader Yasser Arafat says that he believes those was counsel for the bombing case and abroad.

Jewish-Arab relations

Middle East peace negotiations

Middle East

Palestinian self-rule areas

Israel

Politics and government

Arafat, Yasir

Palestinian Arabs

AT&T Radio News Retrieval

QUERY:

RESULTS - "What is the status of the trade deficit with Japan"

| RANK | PROGRAM | DATE | STORY | SCORE | LENGTH | HITS |
|------|---------------------------|-------|-------|-------|--------|------|
| 1 | NPR All Things Considered | 05/31 | 3 | 15.63 | 27.65 | 6 |
| 2 | NPR All Things Considered | 05/10 | 15 | 13.89 | 512.12 | 16 |
| 3 | NPR/PRI Marketplace | 06/14 | 1 | 13.82 | 166.10 | 14 |
| 4 | ABC World News Now | 06/13 | 6 | 13.41 | 30.00 | 3 |
| 5 | NPR All Things Considered | 05/21 | 1 | 11.14 | 13.62 | 3 |
| 6 | NPR All Things Considered | 05/31 | 3 | 10.92 | 17.02 | 3 |
| 7 | NPR/PRI Marketplace | 06/14 | 3 | 10.87 | 30.00 | 1 |
| 8 | CNN Headline News | 06/07 | 18 | 9.83 | 183.55 | 6 |
| 9 | NPR/PRI Marketplace | 08/11 | 23 | 9.82 | 203.21 | 11 |
| 10 | NPR/PRI Marketplace | 08/14 | 6 | 9.41 | 90.33 | 4 |

OVERVIEW - NPR All Things Considered 05/10

deficit

status

japan

trade

ASR TRANSCRIPTS - NPR All Things Considered 05/10

"expanding defense cooperation span is a part of our pacific democracy defense program will strengthen are lines and serve on mutual interest that while president clinton is earth credit for renewing inspecting those ties on his recent trip the administration's amateurs and in a factory posturing on **trade** disputes"

"buster and those ties and assess state of the president's recent attempt of damage control in nineteen ninety four that lead administration for both a **trade** war and lost and then declared victory even though present but received nothing the clinton a station shows tunk war dead and then contradictory tactics"

"did not work for the force camp and saving deregulation competition and economic reform the result has been an increase in both the bilateral **trade deficit** and **japanese trade** nationalism the merchandise **trade** that has no sacred is anthony here no but i do not agree with president clinton's decision"

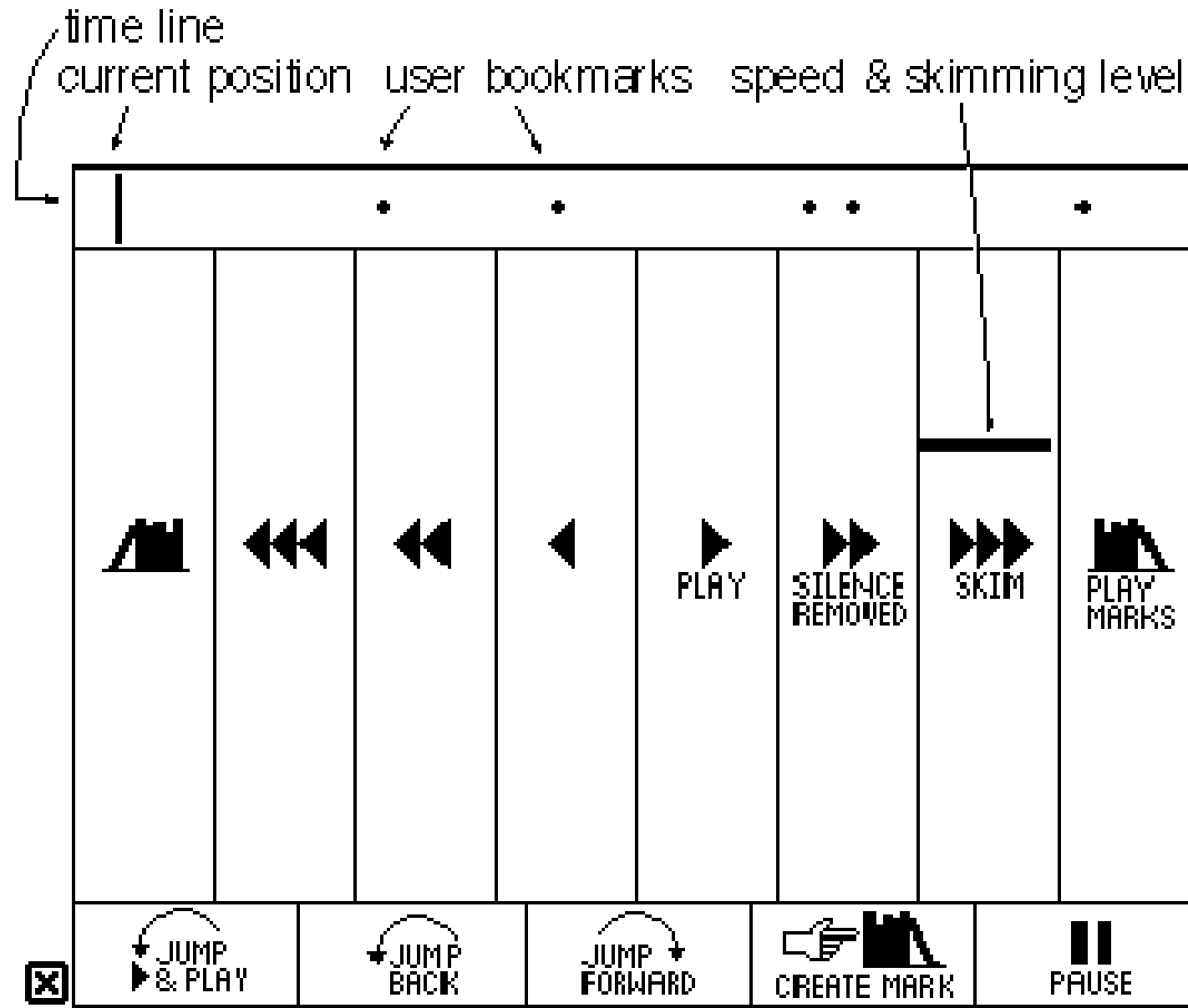
"the normal eyes relations with vietnam until they could could have and should receive more returned from vietnam the decision has been made the case is not closed there are many outstanding issues in our relationship with vietnam was shared economic and other enters can only be realized"

"after the outcome achieved fullest possible accounting for a missing servicemen and vietnam must understand that further progress on the field of the a. m. i. a. issue remain are biased bilateral priority now it is simply that i think we all saw to be very forthright flat out but i have fun"

"that out neo from about are commercial relations with china was incredible is right the nineteen ninety four when a funny decided extension of most favored nation **status** was the best way to promote are long term interest in china"

Selection Length: seconds

MIT “Speech Skimmer”



Comparison with Text Retrieval

- Detection is harder
 - Speech recognition errors
- Selection is harder
 - Date and time are not very informative
- Examination is harder
 - Linear medium is hard to browse
 - Arbitrary segments produce unnatural breaks

A Richer View of Speech

- Speaker identification
 - Known speaker and “more like this” searches
 - Gender detection for search and browsing
- Topic segmentation
 - Vocabulary shift, cue words
 - More natural breakpoints for browsing
- Speaker segmentation
 - Visualize turn-taking behavior for browsing
 - Classify turn-taking patterns for searching

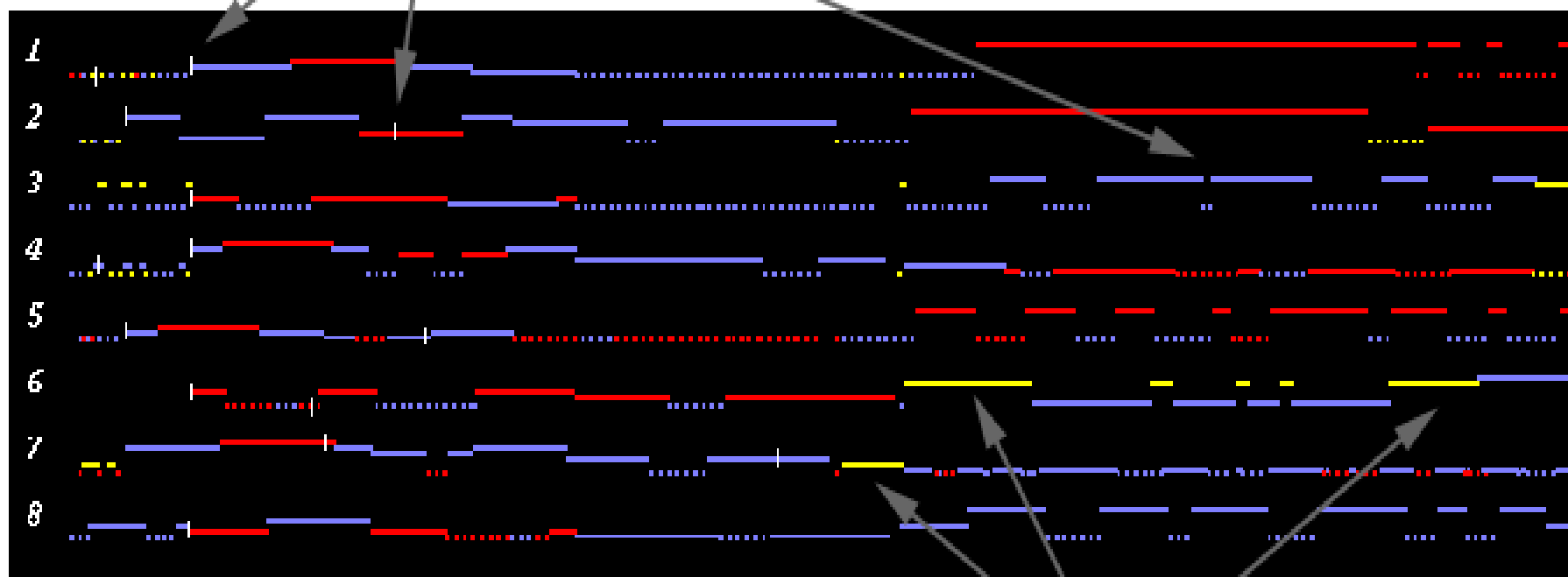
Speaker Identification

- Gender
 - Classify speakers as male or female
- Identity
 - Detect speech samples from same speaker
 - To assign a name, need a known training sample
- Speaker segmentation
 - Identify speaker changes
 - Count number of speakers

Visualizing Turn-Taking

Query Term Location

8 Recordings



Known Speaker

Broadcast News Retrieval Study

- NPR Online
 - Manually prepared transcripts
 - Human cataloging
- SpeechBot
 - Automatic Speech Recognition
 - Automatic indexing

Your search for *"war on terrorism"* matched **162** out of **108945** documents.

Page **1** of **17**

1 [William Bennett](#)

05-08-02 Score: 0.80

Guests: William Bennett *Co-Director, Empower America and Chairman of K12 (an internet-based elementary and secondary school) *Author, *Why We Fight: Moral Clarity and the War on Terrorism* (Doubleday, 2002)<b

2 [Mood In Israel](#)

05-07-02 Score: 0.77

It has been 30 days since Israel entered the West Bank. More than 20 soldiers were killed in what Israeli officials call their own "war on terrorism." NPR's Linda Gradstein asks how the average Israeli feels about the operation and the extent to which it

3 [Voices of Dissent](#)

04-23-02 Score: 0.77

Tavis Smiley speaks with two young editors, Jeremy Glick,

[Tapes & Transcript](#)

Order Online:

[Secure Order Form](#)

[Search for a transcript](#)

Order by Phone:

1.877.NPR.TEXT
(1.877.677.8398)

[Most requested](#)

NPR Series Specials

[Science & Health](#)

[American History](#)

[Religion](#)

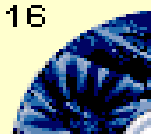
[Foreign Affairs](#)

[Entertainment & Culture](#)

Take the tapes & transcript [survey](#)

all songs considered

BUY THE 16
SONG CD
TODAY!



This segment is from
Talk of the Nation
Wednesday, May 08, 2002

[Listen to Segment](#)

[Listen to Entire Show](#)

[View Program Segments](#)

William Bennett

Guests:

William Bennett

*Co-Director, *Empower America* and Chairman of K12 (an internet-based elementary and secondary school)

*Author, *Why We Fight: Moral Clarity and the War on Terrorism* (Doubleday, 2002)

Matthew Jelacic

*Partner in the Architectural and Industrial Design Firm of Gans and Jelacic

*Joins Neal Conan to discuss the repurposing of "flop homes" -- temporary housing structures for the homeless -- into housing for refugees.

In this hour of *Talk of the Nation*, Neal Conan talks with William Bennett, former education secretary under President Ronald Reagan and drug czar under President George Bush. Bennett will talk about his new book -- *Why We Fight: Moral Clarity and the War on Terrorism* -- and about how he thinks the current President Bush is doing.

SpeechBot

[Simple Search](#)[Power Search](#)[Help](#)[» FAQ](#)[» About SpeechBot](#)[» Feedback](#)

Search for:

Topics:

Dates:

Tip: Try searching a particular topic instead of "*All Topics*"

Search Result: **200 matches** for your querySort results by: **Website****Date****Extract from Transcript***(Transcripts based on [speech recognition](#) are not exact)***PLAY**

extract

The Diane Rehm Show

Mar 14, 2000

...new standards for growing and processing **organic food** the proposal incorporates recommendations that consumer groups and **organic** farmers...[Show me more](#)**PLAY**

extract

PBS Online NewsHour

Dec 21, 2000

...advantage of the exploding demand for **organic** products glickman also said the standards would make things a lot clearer for consumers 1 northern virginia **organic food** shopper...[Show me more](#)**PLAY****Public Interest**

Aug 28, 2000

...get us on the right direction and but you move for a new now we can support farmers markets weekend the point

20 extracts from The Diane Rehm Show - Mar 14, 2000 match your search:
organic food

0 min  51 min

You are here
extract 1 of 20

NEXT
EXTRACT



PLAY
extract

...waga in new and marching tune and diane rehm you as to prevent it the agriculture is proposed new standards for growing and processing **organic food** the proposal incorporates recommendations that consumer groups and **organic** farmers but some say that standard to comply with popular opinion rather than scientific research joining me to discuss **organic food** standards can claim they're against chief marketing standards in and the state years for the u. s. department of agriculture tell a d. on stand and ...

Display of transcript

Extracts from this transcript in order of relevance:

standards for growing and processing organic food the proposal incorporates



PLAY
51 min

The Diane Rehm Show - Mar 14, 2000

[Visit this website](#)

[Search all transcripts from this website](#)

Study Design

- Seminar on visual and sound materials
 - Recruited 5 students
- After training, we provided 2 topics
 - 3 searched NPR Online, 2 searched SpeechBot
- All then tried both systems with a 3rd topic
 - Each choosing their own topic
- Rich data collection
 - Observation, think aloud, semi-structured interview
- Model-guided inductive analysis
 - Coded to the model with QSR NVivo

Criterion-Attribute Framework

| Relevance Criteria | Associated Attributes | |
|--------------------|--|---|
| | NPR Online | SpeechBot |
| Topicality | Story title Brief summary Audio Detailed summary Speaker name | Detailed summary Brief summary Audio Highlighted terms |
| Story Type | Audio Detailed summary Short summary Story title Program title | Audio Program title |
| Authority | Speaker name Speaker's affiliation | |

Lessons Learned from the MALACH Project

Applying New Technologies to Improve Intellectual Access
to Large Oral History Collections

Douglas W. Oard

College of Information Studies and
Institute for Advanced Computer Studies



ZÁPADOČESKÁ
UNIVERZITA
V PLZNI



Outline

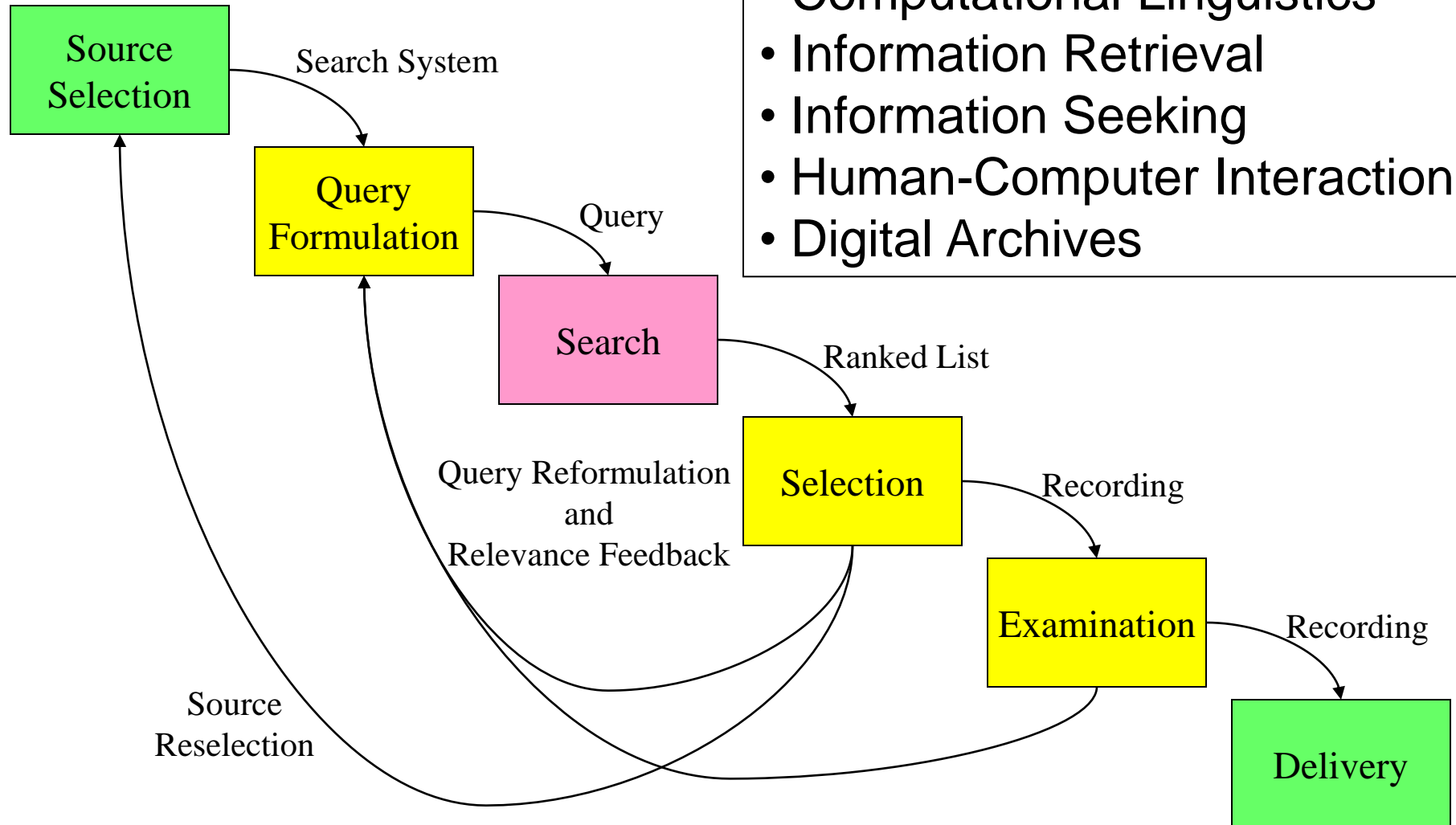
- The MALACH Project
- Learning from our users
- Evaluating new access technologies
- What's left to do?

Indexing Options

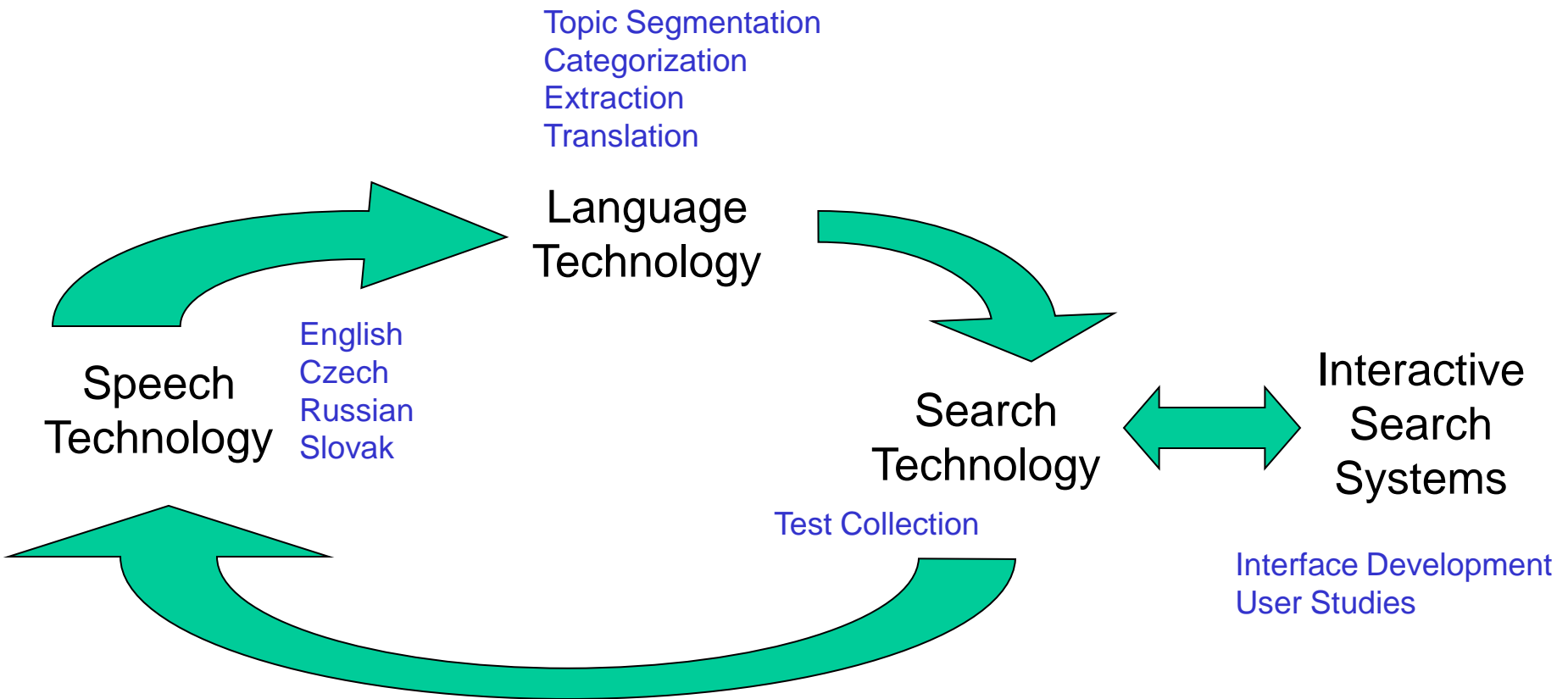
- Transcript-based (e.g., NASA)
 - Manual transcription, editing by interviewee
- Thesaurus-based (e.g., Shoah Foundation)
 - Manually assign descriptors to points in interview
- Catalog-based (e.g., British Library)
 - Catalog record created from interviewer's notes
- Speech-based (MALACH)
 - Create access points with speech processing

Supporting the Search Process

- Speech Processing
- Computational Linguistics
- Information Retrieval
- Information Seeking
- Human-Computer Interaction
- Digital Archives



The MALACH Project



The MALACH/CLEF Team

USA

- USC (Shoah Foundation)
 - Sam Gustman
- IBM TJ Watson
 - Bhuvana Ramabhadran
 - Martin Franz
- U. Maryland
 - Doug Oard
 - Dagobert Soergel
- Johns Hopkins
 - Zak Schafran

Asia

- IBM India
 - Nanda Kambhatha

Europe

- U. Cambridge (UK)
 - Bill Byrne
- Charles University (CZ)
 - Jan Hajic
 - Pavel Pecina
- U. West Bohemia (CZ)
 - Josef Psutka
 - Pavel Ircing
- Dublin City University (IE)
 - Gareth Jones
- Budapest U Tech+Econ (HU)
 - Tibor Fegyo
 - Peter Mihajlik

- Spontaneous conversational speech
- Digitized
- Large
 - 116,000 hours; 52,000 people; 32 languages
- Full-description segment indexing
 - 10,000 hours, ~150,000 topical segments
- Real users



Interview Excerpt

- Content characteristics
 - Domain-specific terms
 - Named entities
- Audio characteristics
 - Accented (this one is unusually clear)
 - Two channels (interviewer / interviewee)
- Dialog structure
 - Interviewers have different styles

Outline

- The MALACH Project
 - Learning from our users
- Evaluating new access technologies
- What's left to do?

Who Uses the Collection?

Discipline

- History
- Linguistics
- Journalism
- Material culture
- Education
- Psychology
- Political science
- Law enforcement

Products

- Book
- Documentary film
- Research paper
- CDROM
- Study guide
- Obituary
- Evidence
- Personal use

2002-2003 Observational Studies

8 independent searchers

- Holocaust studies (2)
- German Studies
- History/Political Science
- Ethnography
- Sociology
- Documentary producer
- High school teacher

8 teamed searchers

- All high school teachers

Thesaurus-based search

Rich data collection

- Intermediary interaction
- Semi-structured interviews
- Observer's notes
- Participant notes
- Screen capture

Qualitative analysis

- Theory-guided coding
- Abductive reasoning



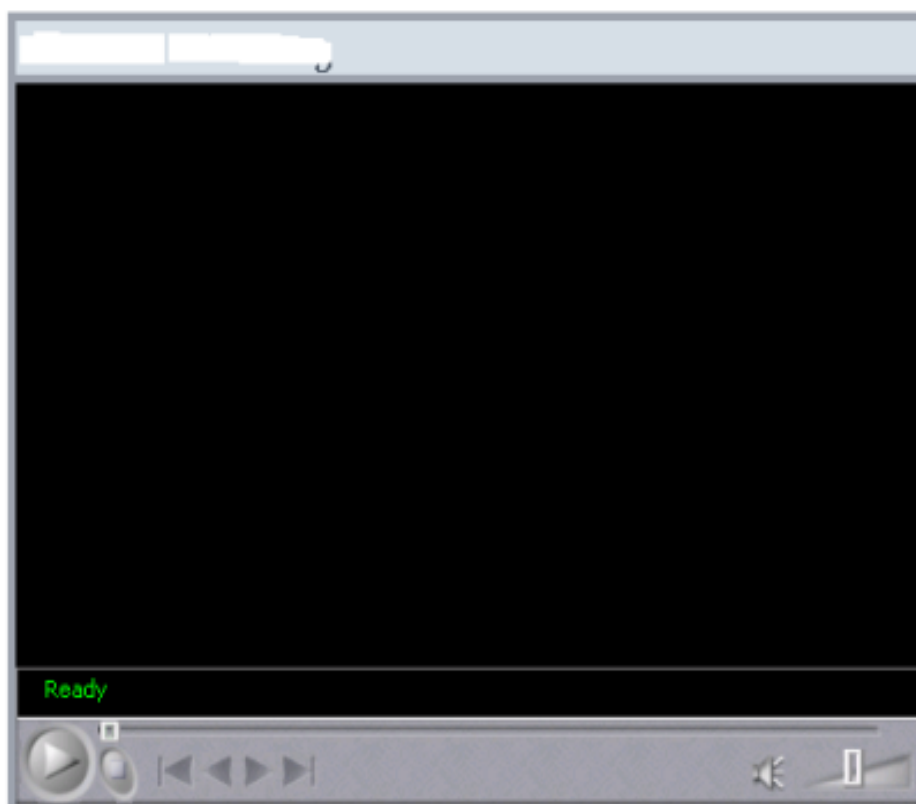
A D V A N C E D S E A R C H

V I E W I N G S C R E E N

Right-click on your mouse for additional video controls.

- ZOOM adjusts the size of the video screen.
- VOLUME adjusts the sound.

To start video, click here:



Choose One:

- ◀ [Video](#)
[Slide Show](#)

Download Time:
less than 15 minutes

All Keywords in this Segment ▶

[Biographical Profile](#)

[Special Thanks](#)

Keywords for [Segment #1](#)

education
family life
Cologne (Germany)
socioeconomic status
divorces and separations
religious holidays, Jewish
Germany 1920 (June 6) - 1923 (August 12)

- [Maximize/Minimize Data](#)
- [Next Segment](#)
- [Previous Segment](#)
- [Back to Search Results](#)
- [Choose New Search](#)
- [Save to My Projects](#)

Scholars: Relevance Criteria

Information
Need

Query

Relevance Judgment

Situation n

S1

Query Reformulation

Query & Browse

- Topicality (76%)
- Comprehensibility (2%)
- Novelty of content (1%)
- Acquaintance (1%)

Browse Only

- Accessibility (6%)
- Richness (6%)
- Emotion (3%)
- Duration (2%)
- Miscellaneous (3%)

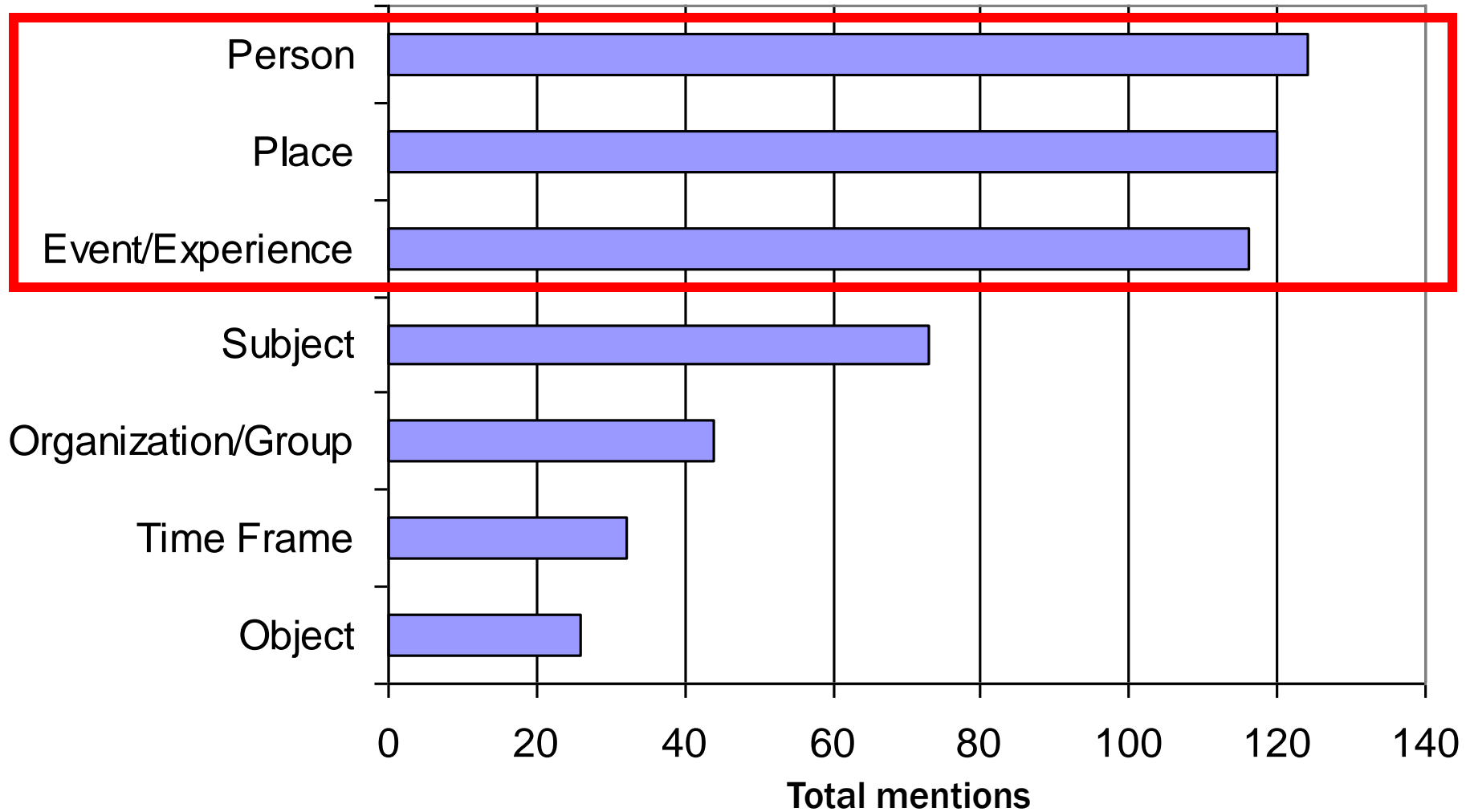
Gap n

G1

Action n

A1

Scholars: Topic Types



6 Scholars, 1 teacher, 1 movie producer, working individually

Basis for Query Reformulation

- Searcher's prior knowledge (31%)
- Viewing an interview or segment (20%)
- Help from an intermediary (15%)
- Thesaurus (13%)
- Assigned descriptors (12%)
- Pre-interview questionnaire (6%)
- Result set size (4%)

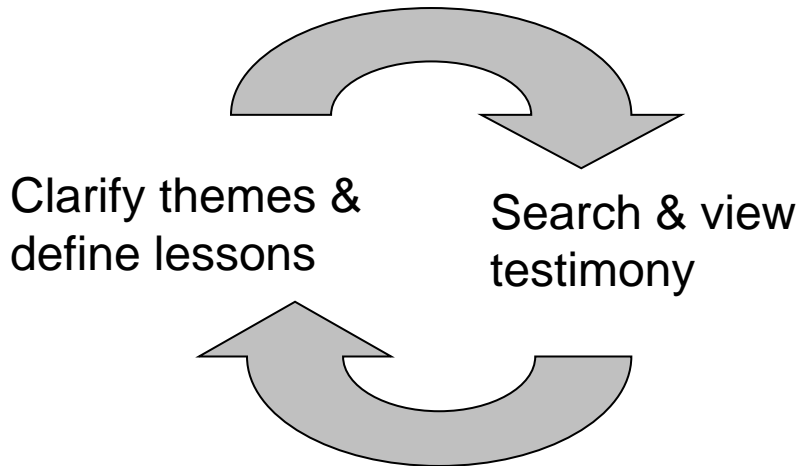
**Total
Of 40%
From
Classification**

Teachers: Search Strategies

- Scenario-guided searching
 - Look for experiences that would reflect themes
- Proxy searching
 - Descriptors & interview questions
- Browsing
 - Listened to testimonies for inspiration
 - Used thesaurus for ideas and orientation

Iterating Searching & Planning

Group discussions clarify themes and define activities, which hone teachers' criteria



Clarify themes & define lessons

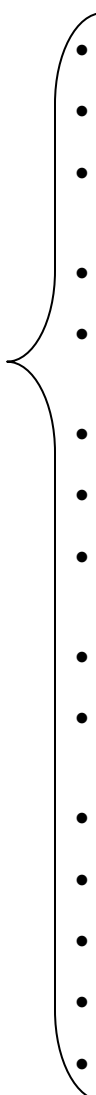
Search & view testimony

Testimonies give teachers ideas on what to discuss in the classroom (topic) and how to introduce it (activity).

“Yesterday in my search, I just felt like I was kind of going around in the dark. But that productive writing session really directed my search.”

“We actually looked at the testimonies that we found before we started writing the lesson ... We really started with the testimony and built our lesson around it.”

Teachers: Relevance Criteria

- Relevant to teaching content/ method
 - A: Relationship to theme
 - B: As part of broader curriculum
 - C: Characteristics of the story
 - D: Relationship of story to student
 - E: Represents different populations
 - F: Characteristics of Oral History
 - Appropriateness
 - Developmental
 - Acceptability to stakeholders
 - Length-to-contribution ratio
 - Technical production quality
- 
- B: Relates to other schoolwork
 - B: Variety for the classroom
 - B: Vocabulary
 - C: Positive Message for Students
 - C: Role of Interviewee in Holocaust Events
 - D: Students connect with passage
 - D: Students identify with interviewee
 - D: Radical difference from students' reality
 - E: Age of interviewee at time of events
 - E: Race
 - F: Expressive power
 - F: Language & verbal expression
 - F: Nonverbal Communication
 - F: Diction
 - F: Flow of interview

Tie Into Curriculum

- Relates to other school work

“I am dealing with the Holocaust through literature. You know, *Night*, *Anne Frank’s diaries* ... I would very much like to see [segments] that specifically relate to various pieces of literature.”



- Suggests:
 - Query by published example
 - Using vocabulary lists to prime ASR

Linking Students to a Story

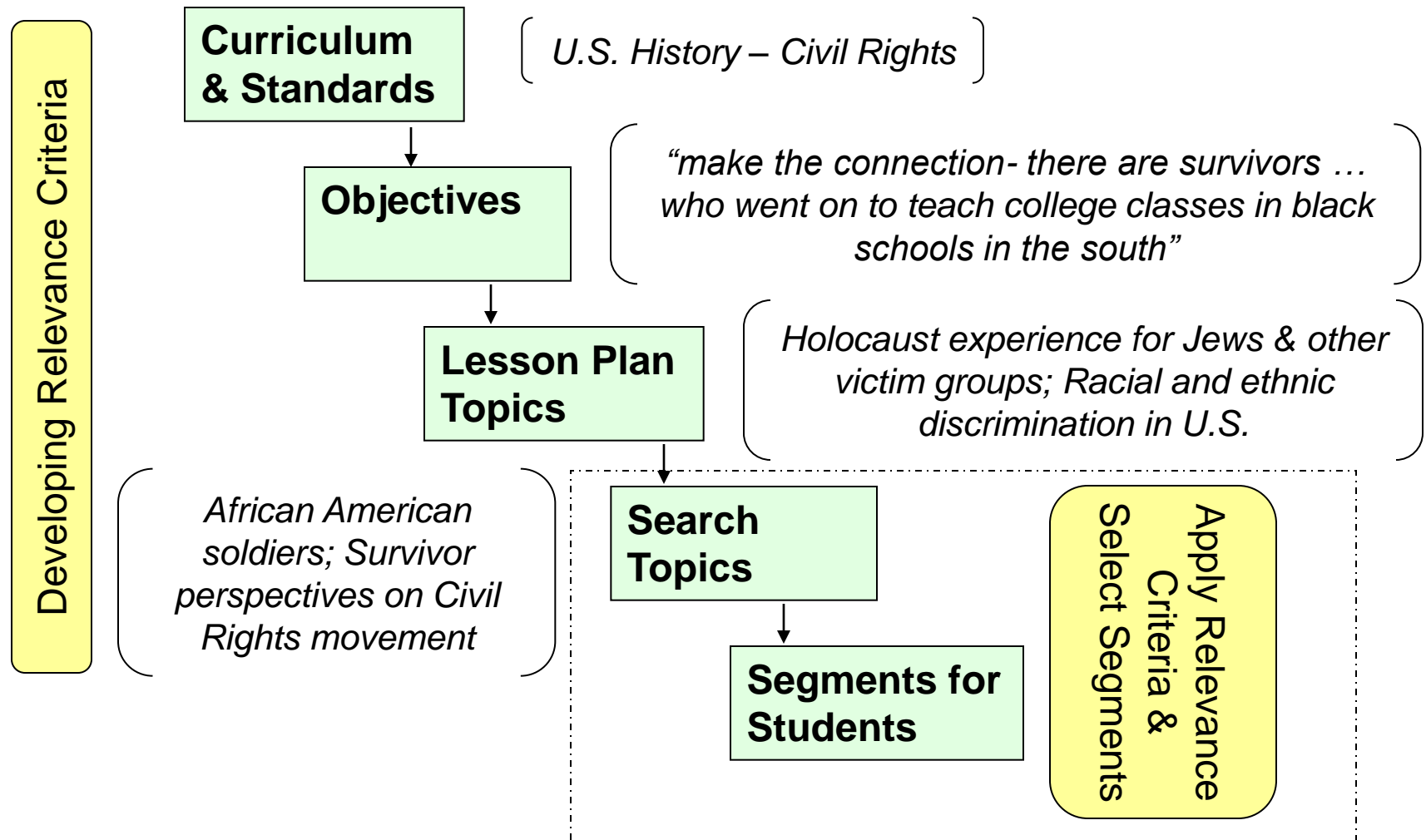
- Students identify with interviewee

“Here she is, this older lady, but she became a 16-year-old girl when she was talking. I saw her in her school uniform on the day of graduation. She was so proud of the way she looked, and she felt so sophisticated.”



- Suggests
 - Demographic search constraints
 - Indexing characteristics, not just words

2006 Observational Study (Teachers)



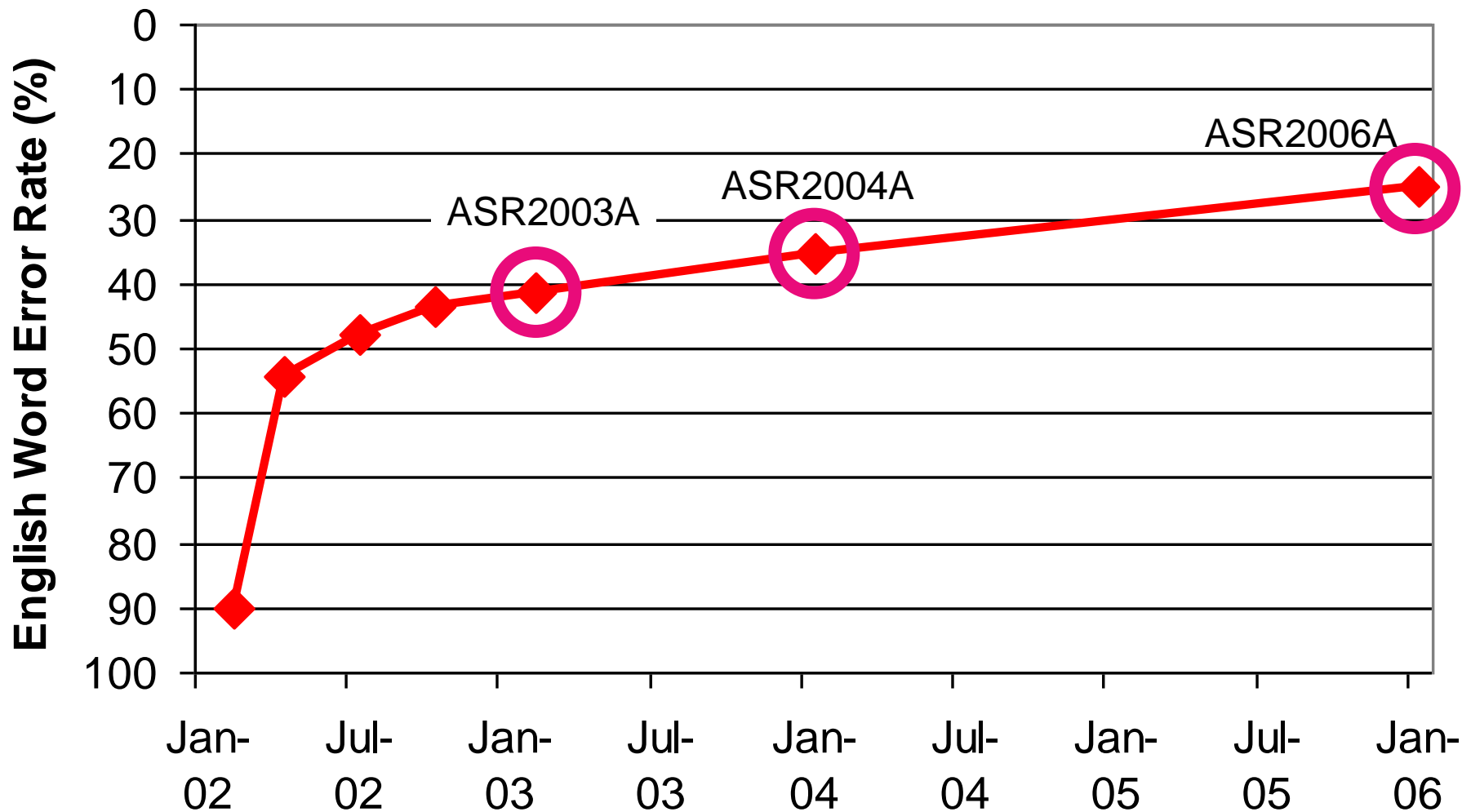
User Study Limitations

- Generalization
 - Exploratory rather than comparative
 - Results influenced by system and domain
- Limited number of study participants
 - Insufficient to observe group differences
- Lack of prior experience with the system
 - Experienced users may behave differently

Outline

- The MALACH Project
- Learning from our users
 - Evaluating new access technologies
- What's left to do?

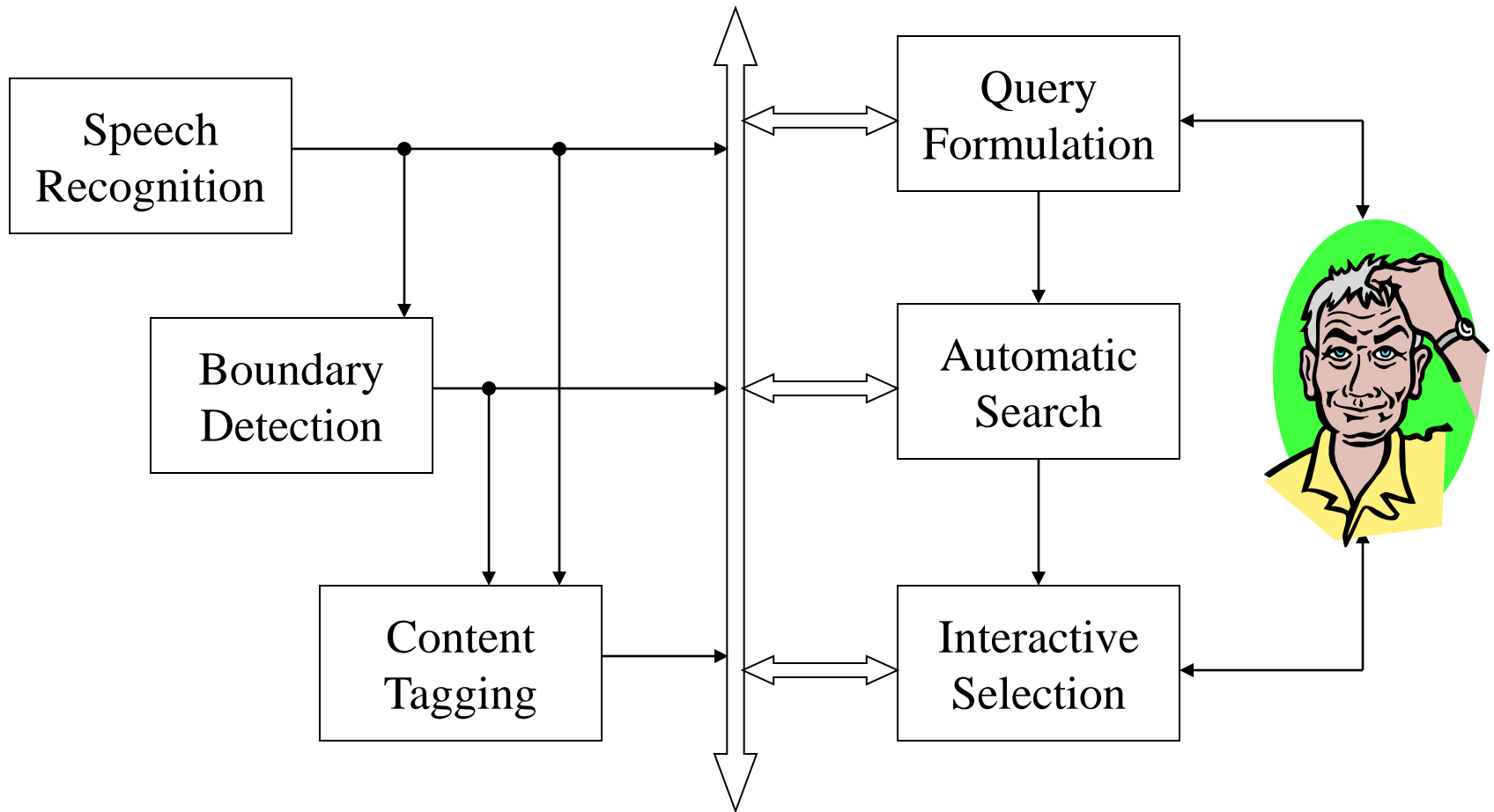
English Transcription Accuracy



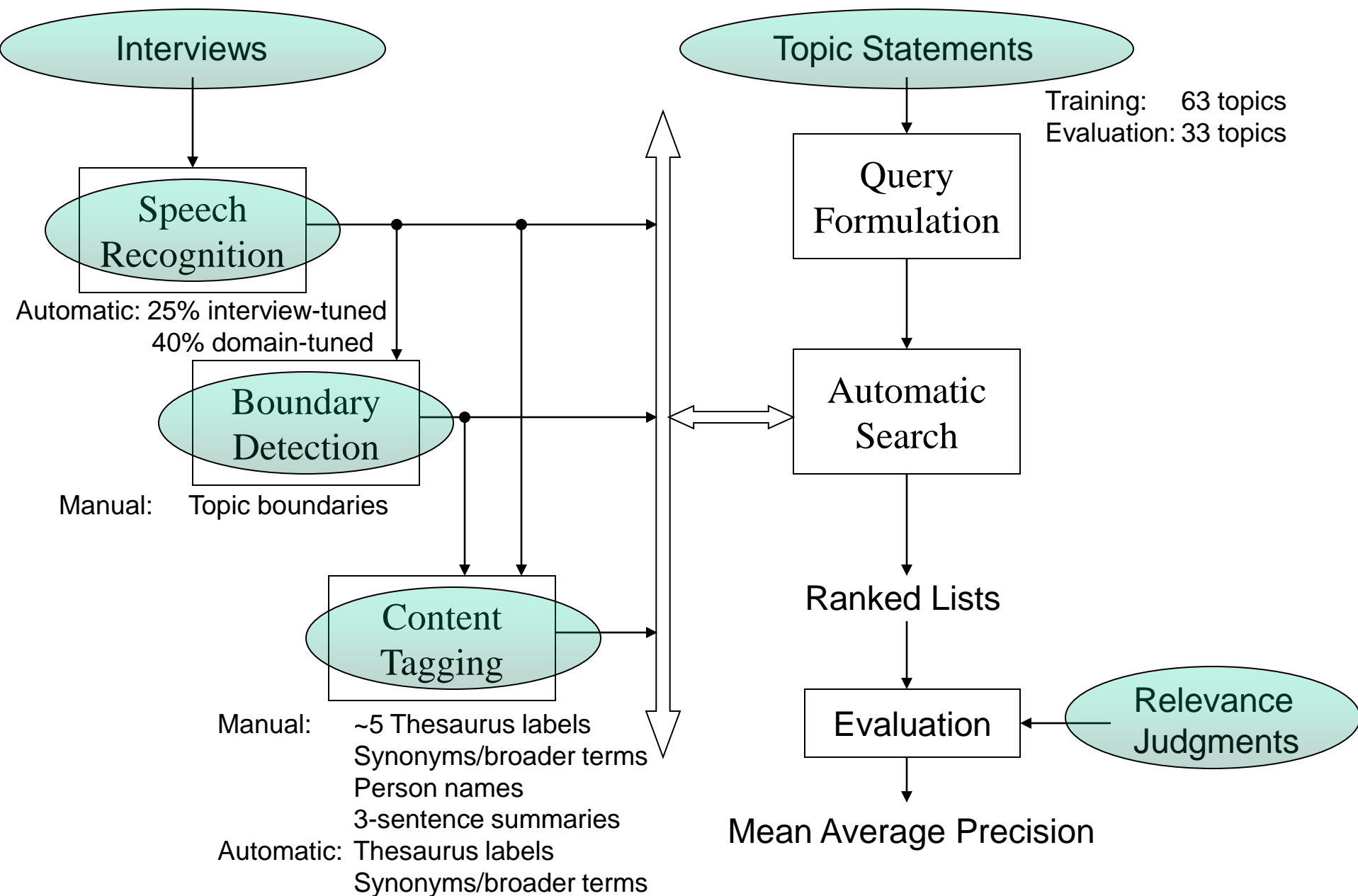
Speech Retrieval Evaluations

- 1996-1998: TREC SDR
 - EN broadcast news / EN queries
- 1997-2004: TDT
 - EN+CN+AR broadcast news / Query by example
- 2003-2004: CLEF CL-SDR
 - EN broadcast news / Many query languages
- 2005-2007: CLEF CL-SR
 - EN+CZ interviews / Many query languages

English Test Collection Design



English Test Collection Design



2006/07 CL-SR English Collection

- 8,104 topically-coherent segments
 - 297 English interviews
 - Known-boundary condition (“Segments”)
 - Average 503 words/segment
- 96 topics
 - Title / Description / Narrative
 - 63 training + 33 evaluation
 - 6 topic languages (CZ, DE, EN, FR, NL, SP)
- Relevance judgments
 - Search-guided + “highly ranked” (pooled)
- Distributed to track participants by ELDA

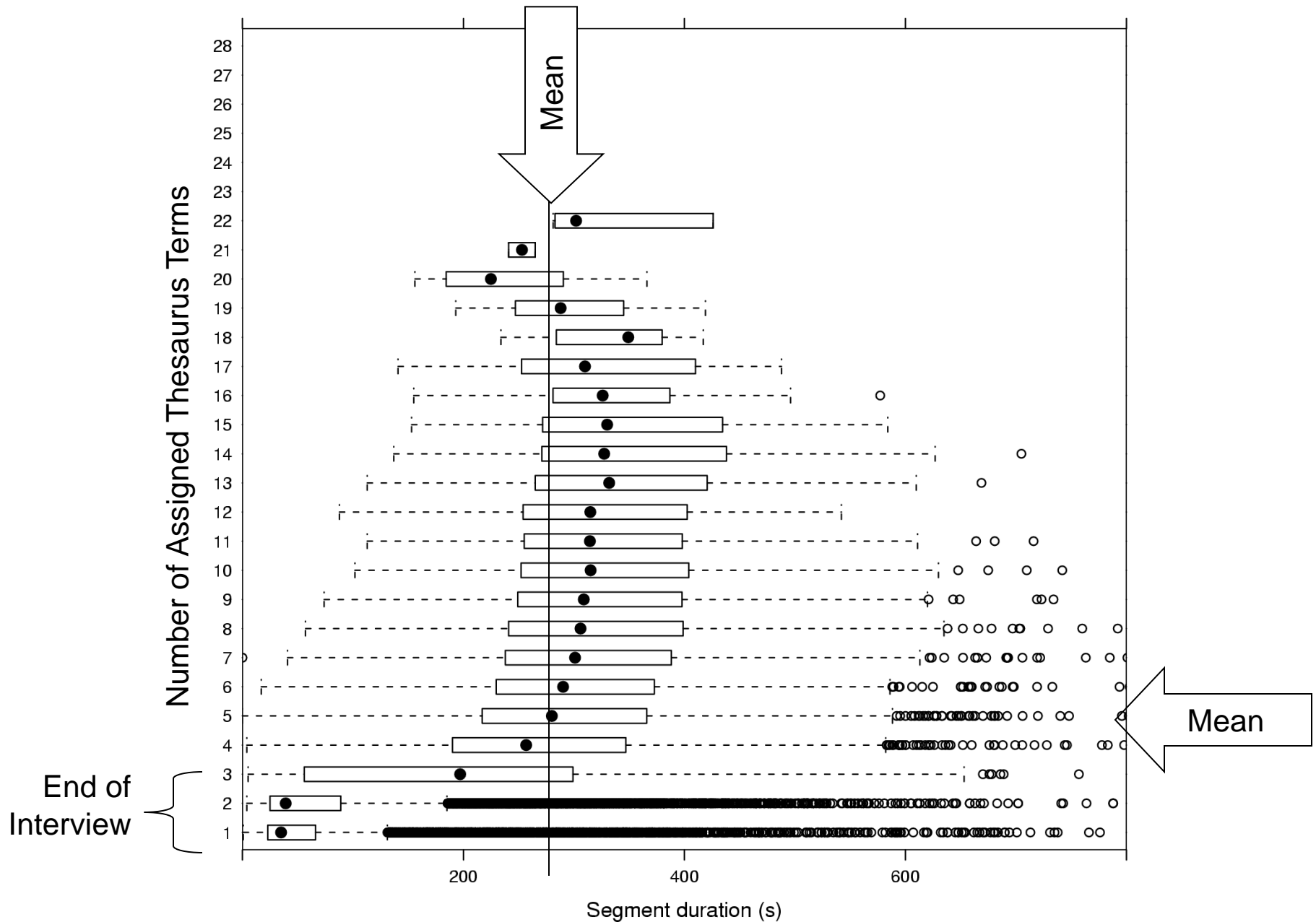
<DOC>
<DOCNO>
<INTERVIEWDATA>
<NAME>
<MANUALKEYWORD>
<SUMMARY>

<ASRTEXT2003A>
<ASRTEXT2004A>
<ASRTEXT2006A>
<ASRTEXT2006B>
<AUTOKEYWORD2004A1>
<AUTOKEYWORD2004A2>

Supplementary Resources

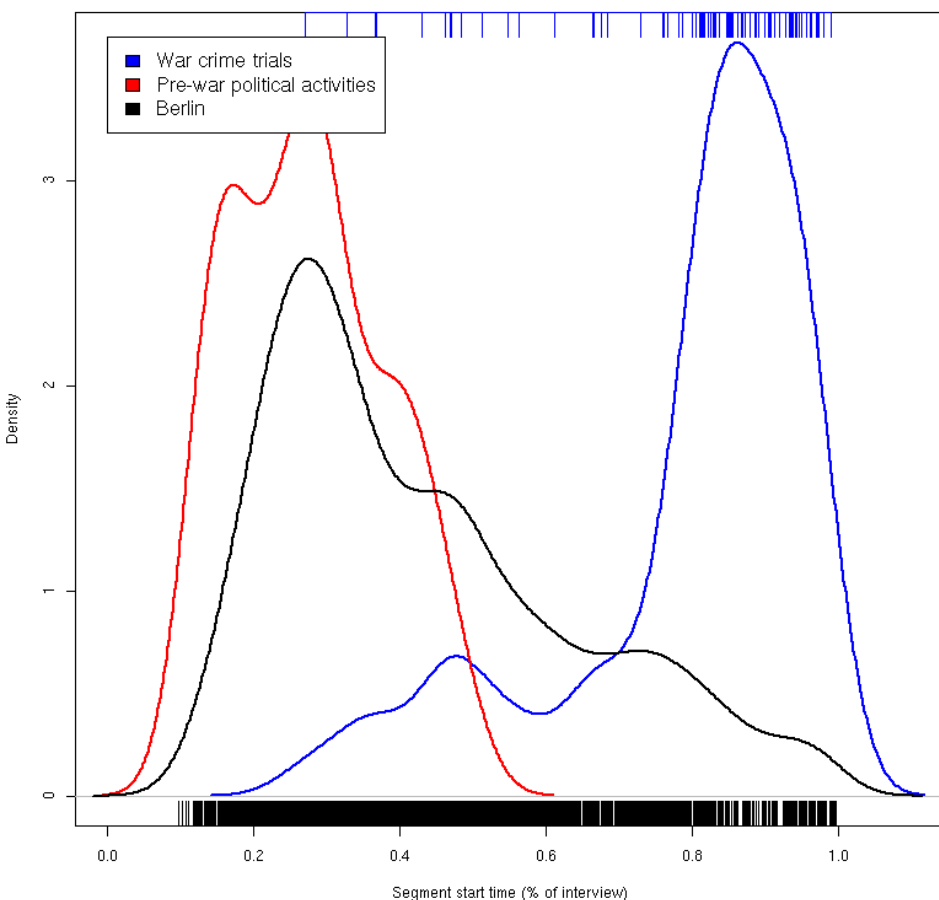
- Thesaurus (Included in the test collection)
 - ~3,000 core concepts
 - Plus alternate vocabulary + standard combinations
 - ~30,000 location-time pairs, with lat/long
 - Is-a, part-whole, “entry vocabulary”
- Digitized speech
 - .mp2 or .mp3
- In-domain expansion collection (MALACH)
 - 186,000 scratchpad + 3-sentence summaries

Manually Assigned Thesaurus Terms



Sequence-Based Classification

Temporal Label Weights (TLW) (Based on Absolute Position)



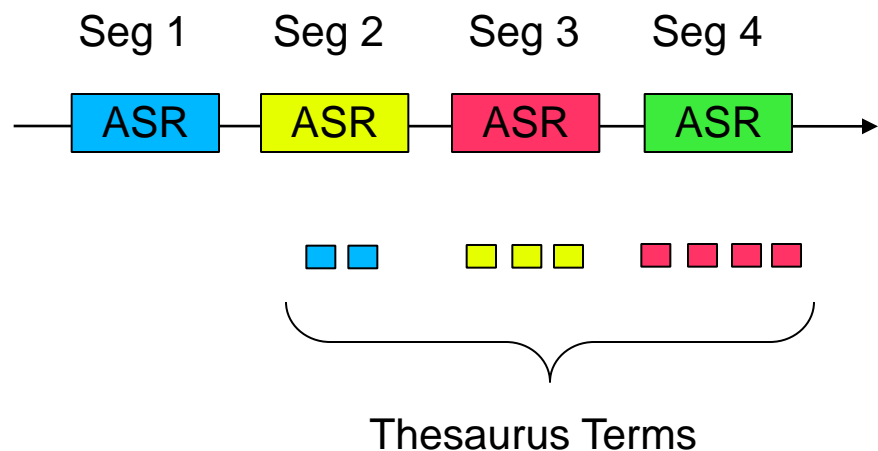
Time-Shifted Classification (TSC) (Based on Relative Position)

Seg 1: "I was born in Berlin ..."

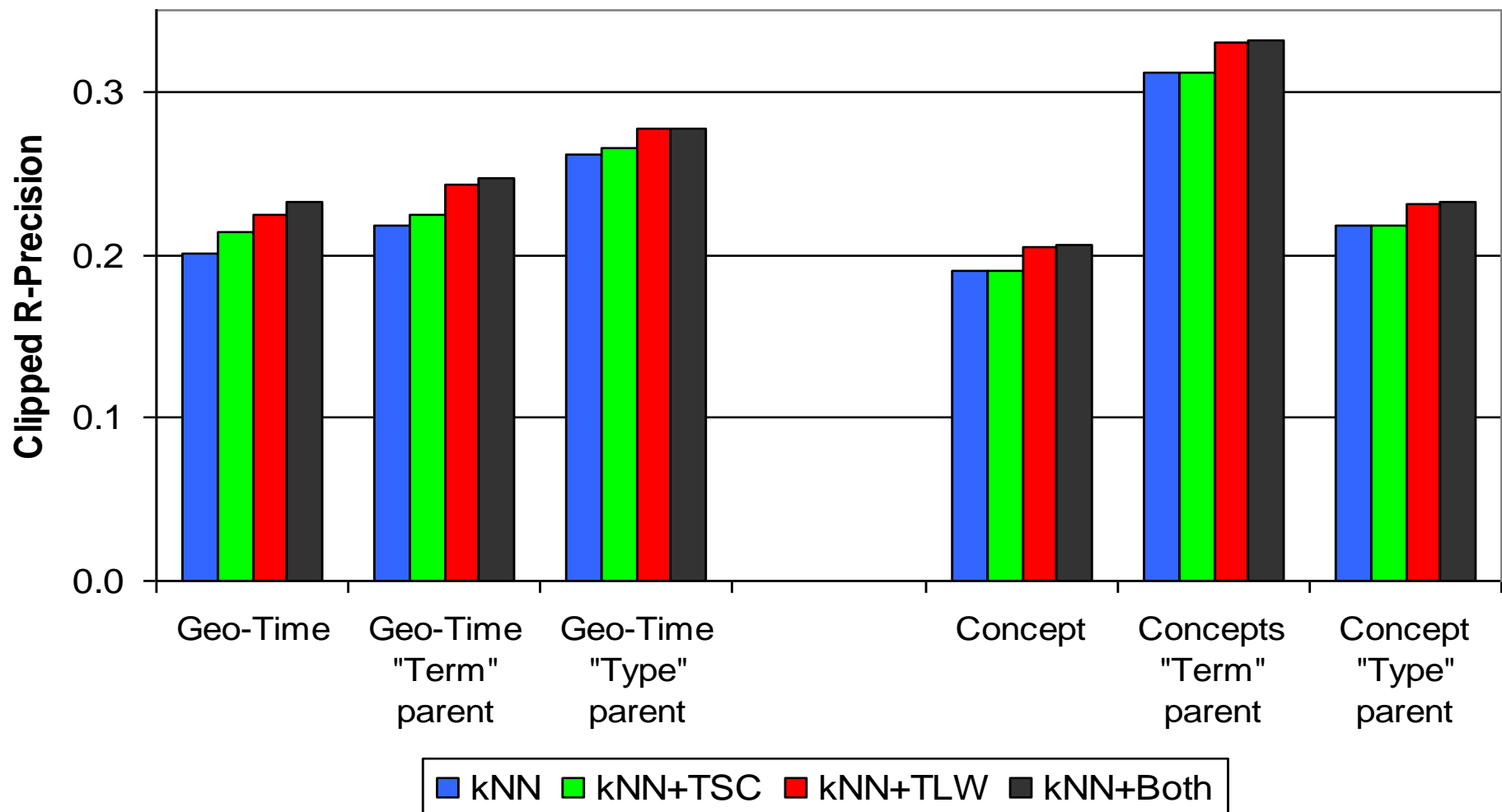
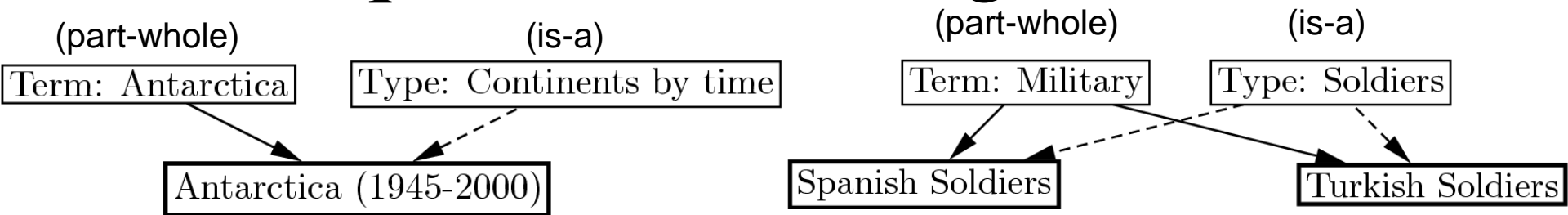
Seg 2: "My playmates included ..."

Seg 3: "Later, we moved to Munich ..."

Seg 4: "That's when things started to ..."



Sequence Modeling Results



An English Topic

Number: 1148

Title: Jewish resistance in Europe

Description:

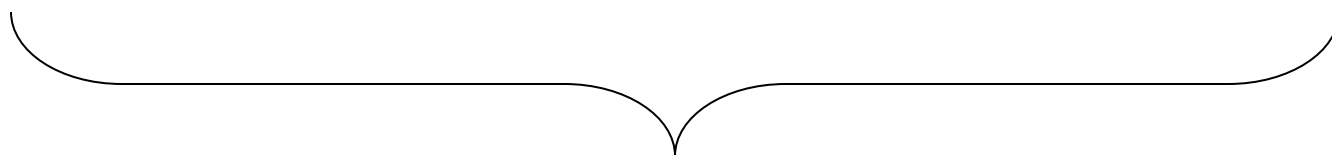
Provide testimonies or describe actions of Jewish resistance in Europe before and during the war.

Narrative:

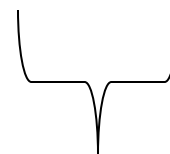
The relevant material should describe actions of only- or mostly Jewish resistance in Europe. Both individual and group-based actions are relevant. Type of actions may include survival (fleeing, hiding, saving children), testifying (alerting the outside world, writing, hiding testimonies), fighting (partisans, uprising, political security) Information about undifferentiated resistance groups is not relevant.

The 118 CL-SR Topics

| | | | | |
|--------------------------------|----------------------------------|-------------------------------------|-------------|-----------------------|
| 38 2005 English Training | 25 2005 English Evaluation | 33 2006/07 English Evaluation | 9 Unused | 13 Czech Safety |
|--------------------------------|----------------------------------|-------------------------------------|-------------|-----------------------|



28
2006 Czech
Evaluation
+
41
2007 Czech
Evaluation



1
2006 Czech
Evaluation
+
3
2007 Czech
Evaluation

5-level Relevance Judgments

Binary qrels

- “Classic” relevance (to “food in Auschwitz”)
 - Direct Knew food was sometimes withheld
 - Indirect Saw undernourished people
- Additional relevance types
 - Context Intensity of manual labor
 - Comparison Food situation in a different camp
 - Pointer Mention of a study on the subject

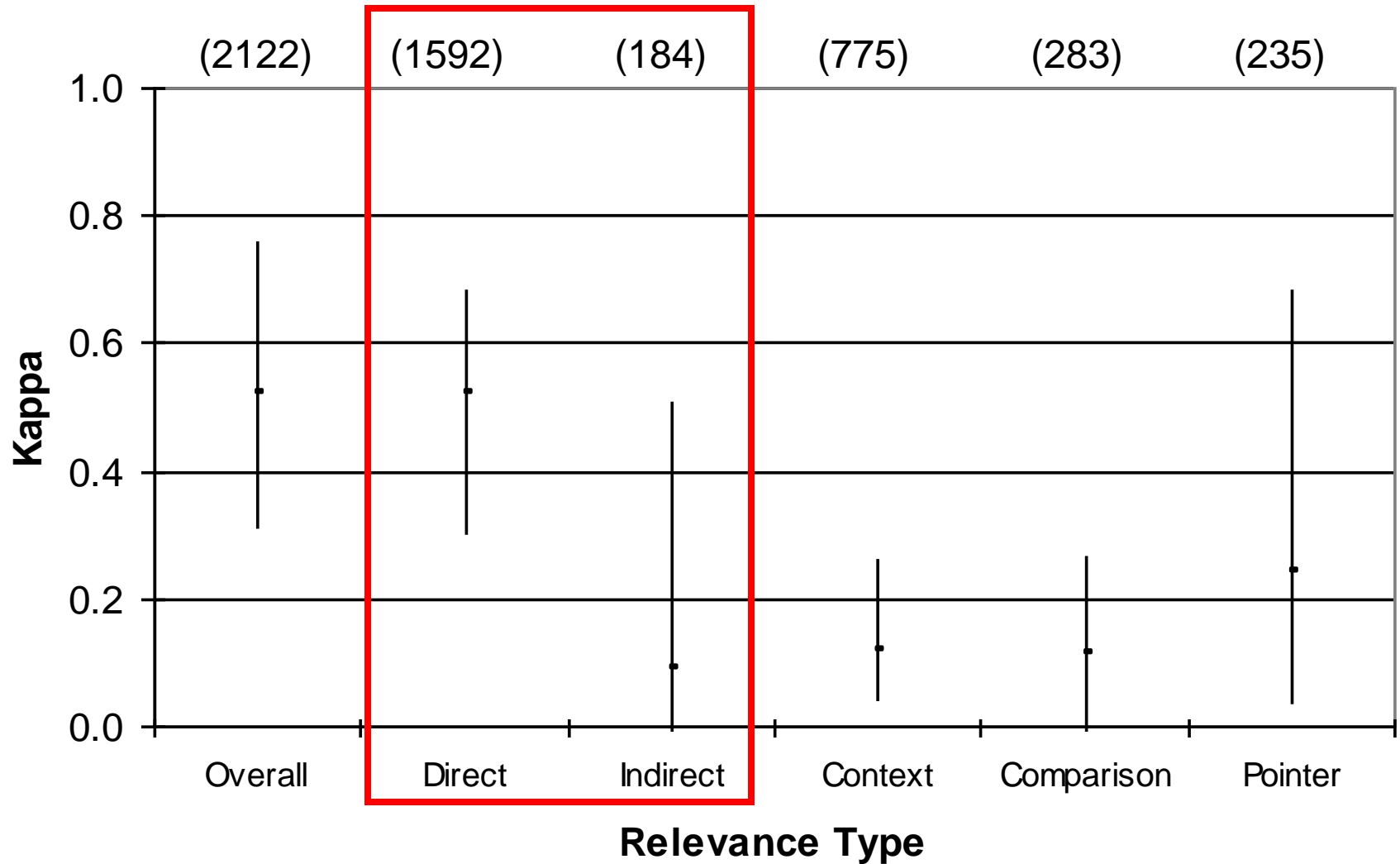
English Assessment Process

- Search-guided
 - Iterate topic research/query formulation/judging
 - Grad students in History, trained by librarians
 - Essential for collection reuse with future ASR
 - Done between topic release and submission
- Highly-ranked (=“Pooled”)
 - Same assessors (usually same individual)
 - Pools formed from two systems per team
 - Top-ranked documents (typically 50)
 - Chosen in order recommended by each team
 - Omitting segments with search-guided judgments

Quality Assurance

- 14 topics independently assessed
 - 0.63 topic-averaged kappa (over all judgments)
 - 44% topic-averaged overlap (relevant judgments)
 - Assessors later met to adjudicate
- 14 topics assessed and then reviewed
 - Decisions of the reviewer were final

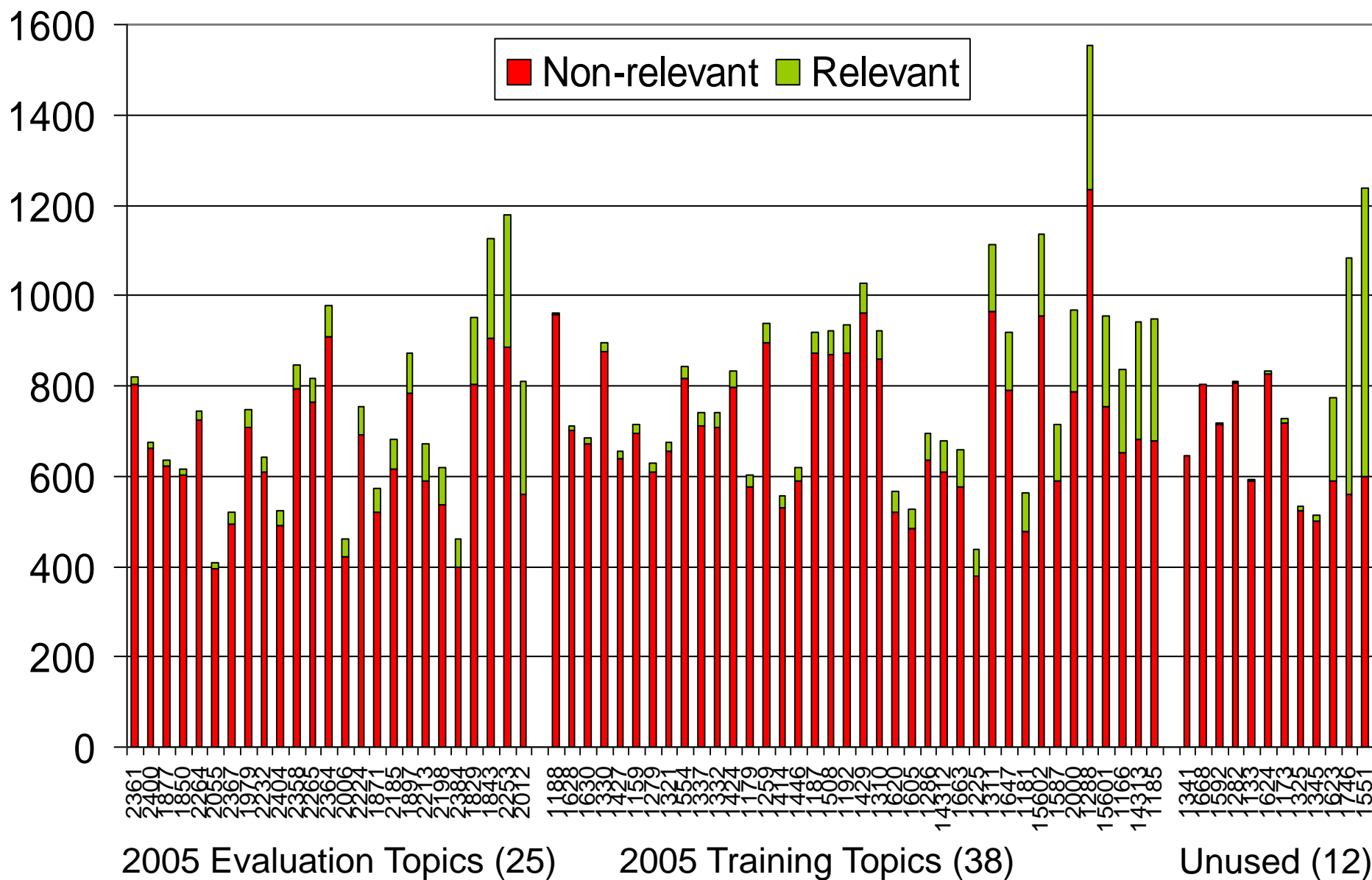
Assessor Agreement (2004)



44% topic-averaged overlap for Direct+Indirect 2/3/4 judgments

14 topics, 4 assessors in 6 pairings, 1806 judgments

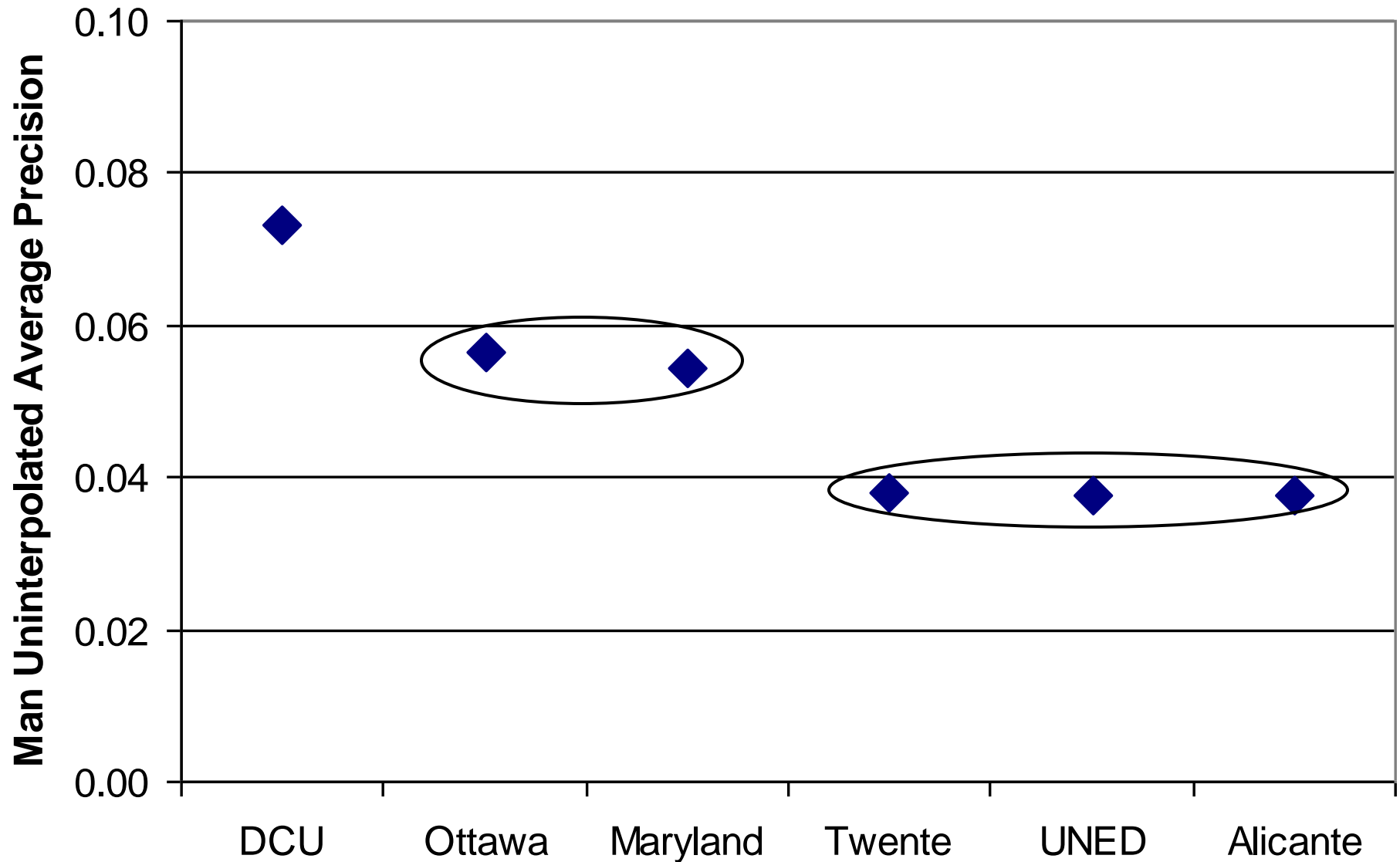
English Relevance Judgments



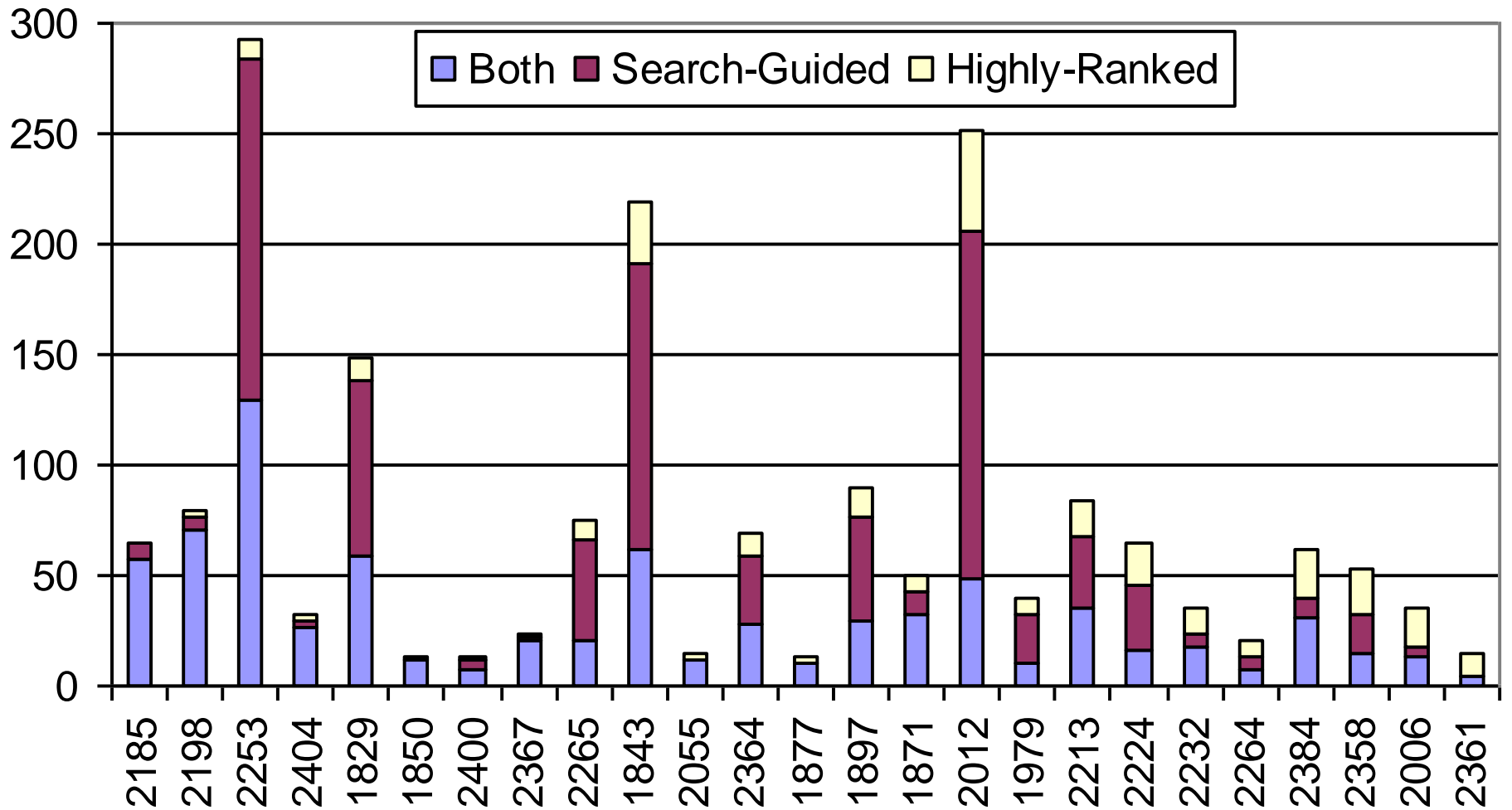
CLEF-2006 CL-SR Overview

- 2 tasks: English segments, Czech start times
 - Max of 5 “official” runs per team per task
 - Baseline English run: ASR / English TD topics
- 7 teams / 6 countries
 - Canada: Ottawa (EN, CZ)
 - Czech Republic: West Bohemia (CZ)
 - Ireland: DCU (EN)
 - Netherlands: Twente (EN)
 - Spain: Alicante (EN), UNED (EN)
 - USA: Maryland (EN, CZ)

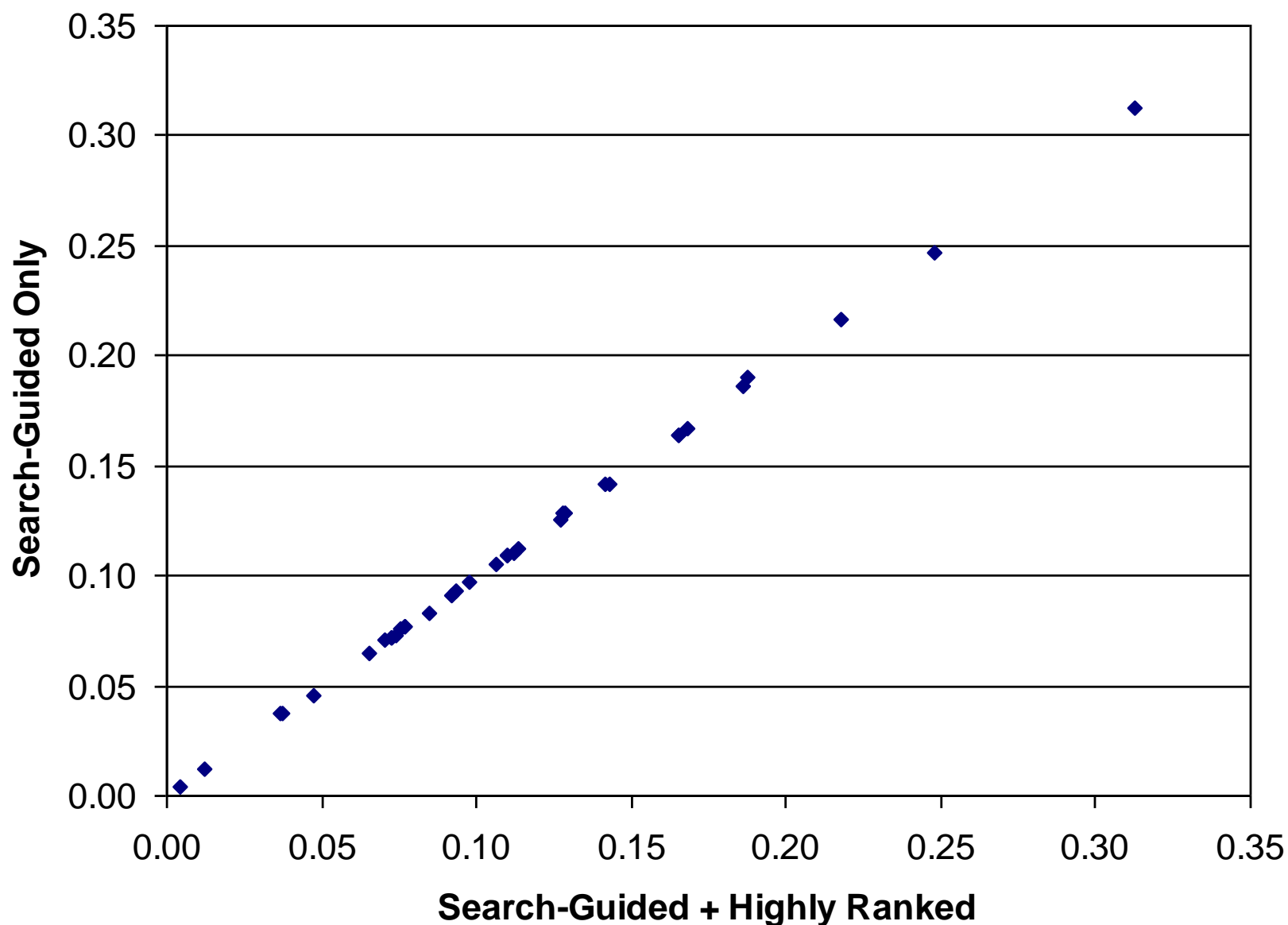
2006 English Results



Contributions by Assessment Type

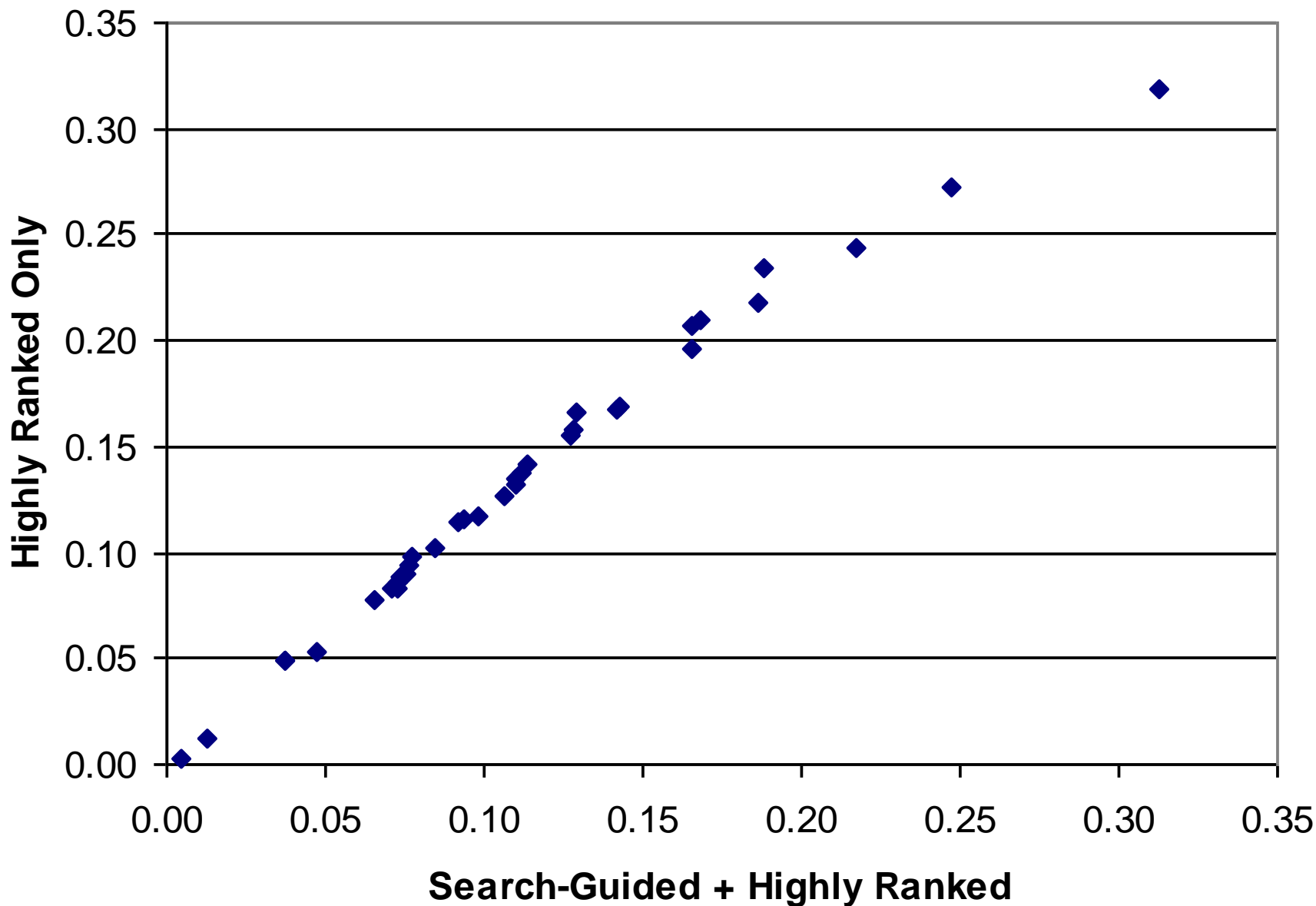


Search-Guided Rankings (2005)



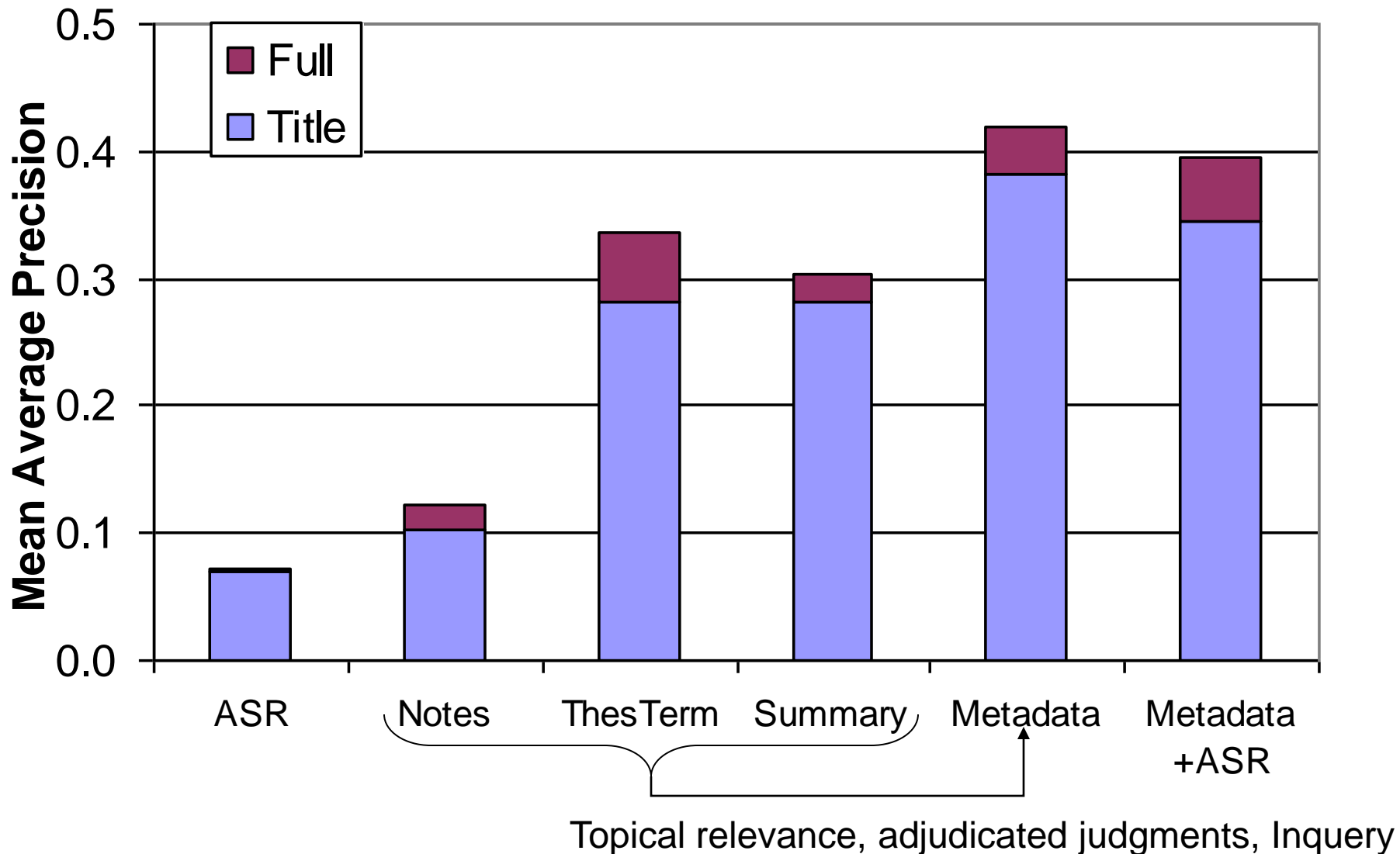
Mean uninterpolated average precision, 35 Official CLEF-2005 runs, 25 evaluation topics

Highly-Ranked Rankings (2005)



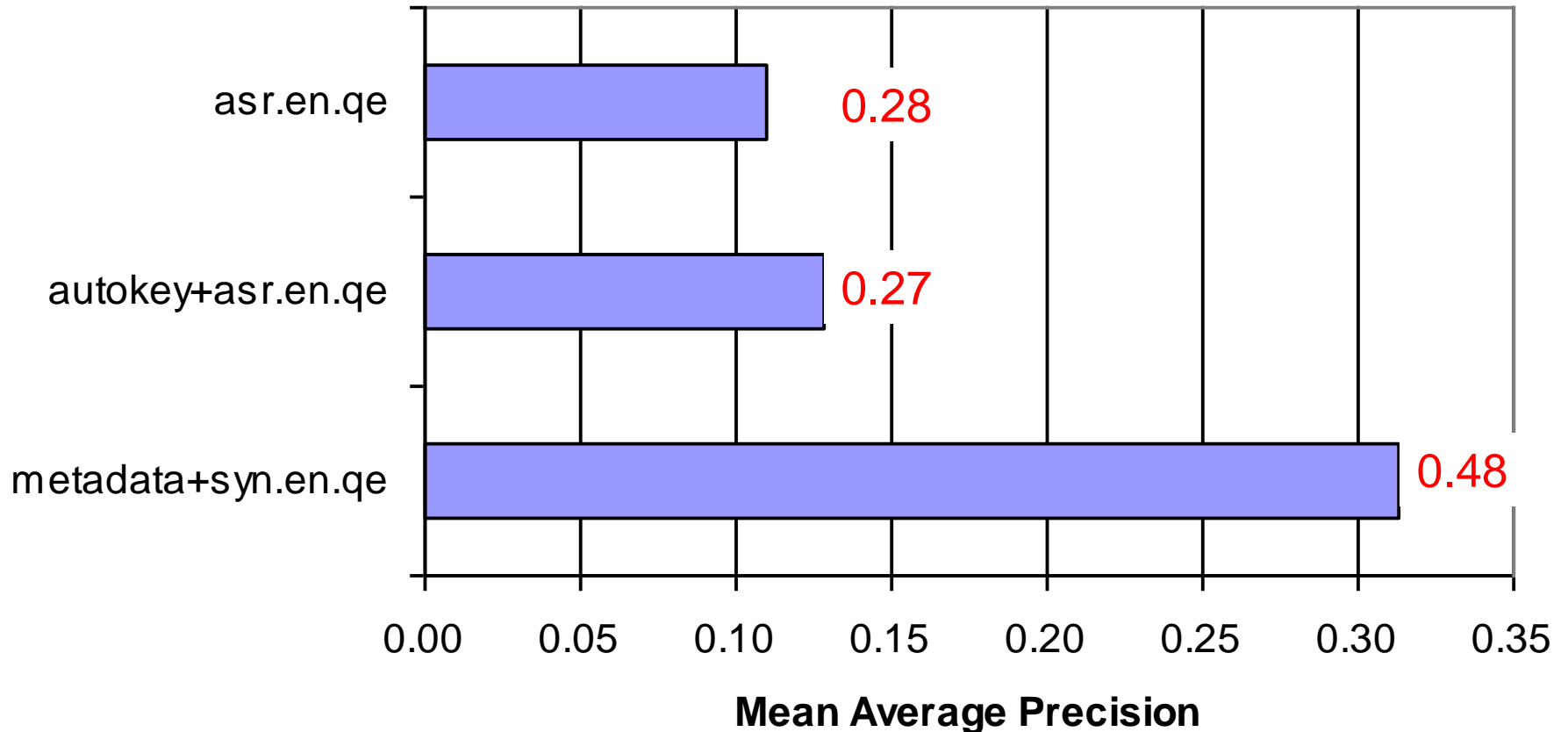
Mean uninterpolated average precision, 35 Official CLEF-2005 runs, 25 evaluation topics

Comparing Index Terms (2003)



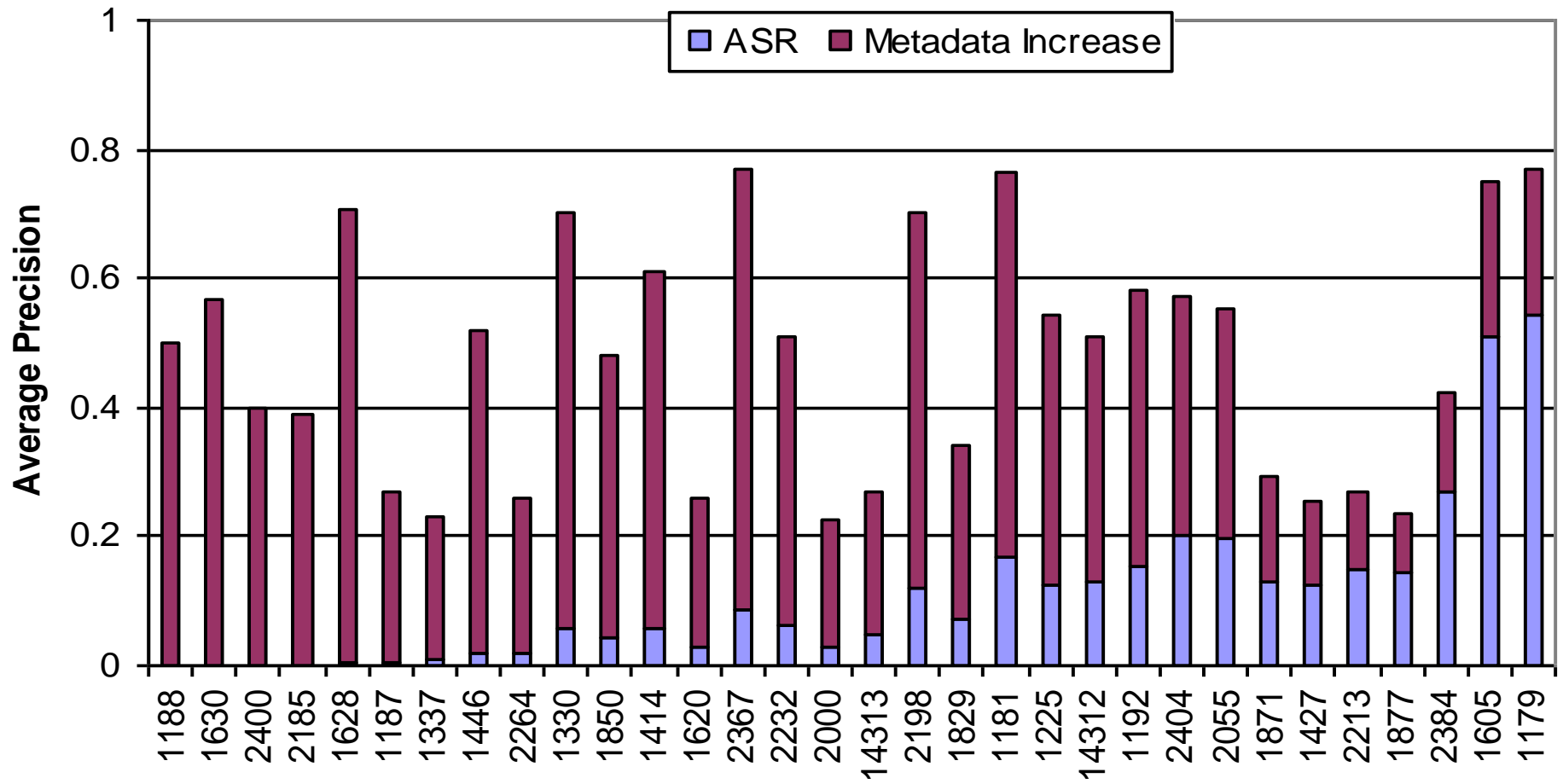
Maryland English Results (2005)

Precision@10



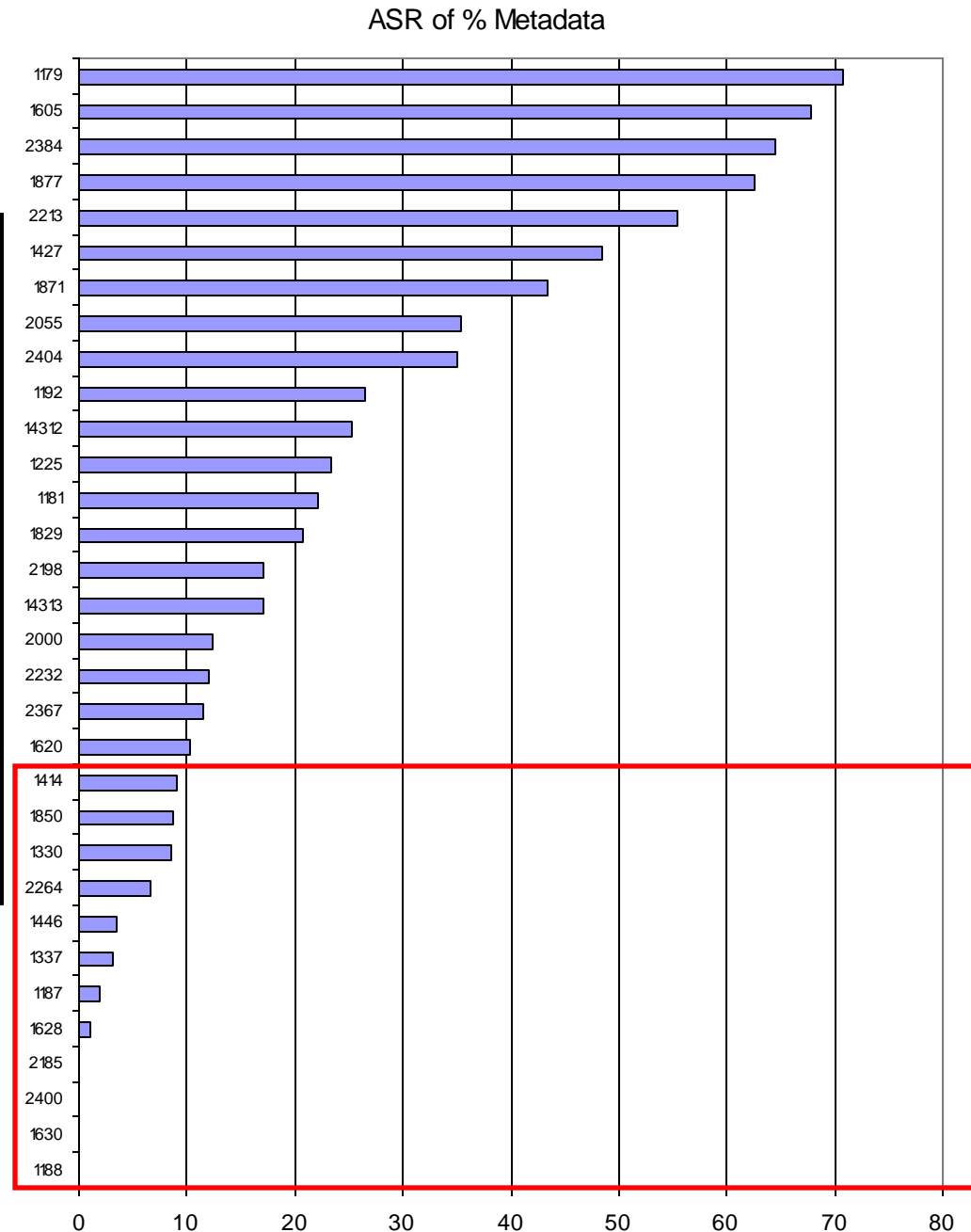
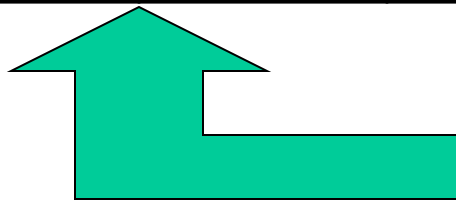
25 evaluation topics, English TD queries

Comparing ASR with Metadata (2005)



Error Analysis (2005)

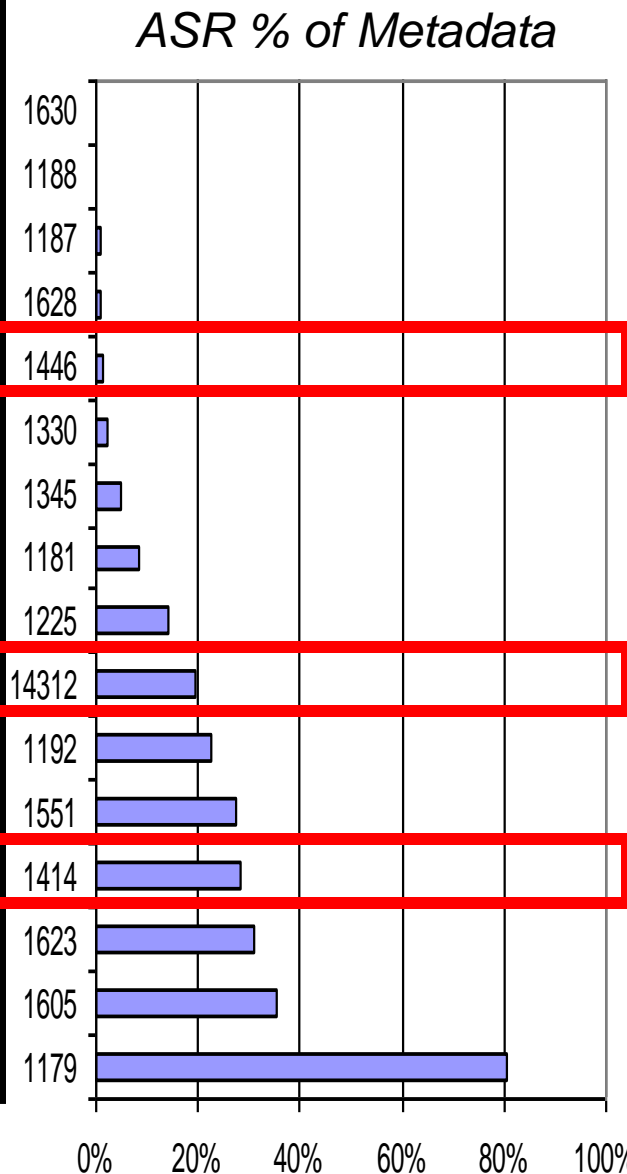
| Somewhere in ASR (ASR/Metadata) | Only in Metadata | |
|------------------------------------|---------------------|-----|
| wallenberg (3/36)* rescue jews | | |
| wallenberg (3/36) | eichmann | |
| abusive female (8/81) personnel | | |
| minsko (21/71) ghetto underground | | |
| art auschwitz | | |
| labor camps | ig farben | |
| slave labor | telefunken | aeg |
| holocaust | sinti roma | |
| sobibor (5/13) death camp | | |
| witness | eichmann | |
| jews | volkswagen | |



CLEF-2005 training + test – (metadata < 0.2), ASR2004A only, Title queries, Inquiry 3.1p1

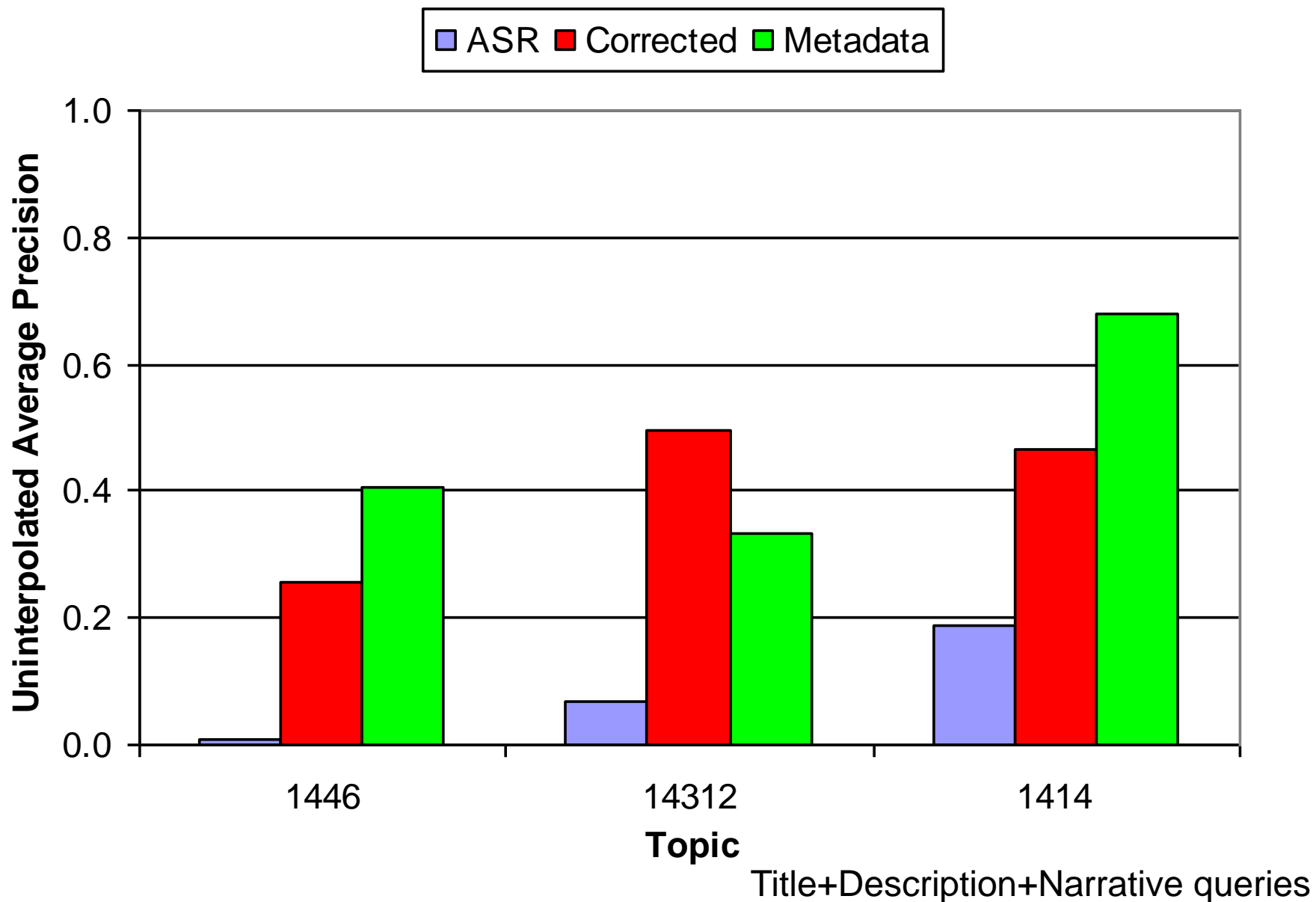
2004 Error Analysis

| <i>Somewhere in ASR Results (bold occur in <35 segments)</i> | <i>in ASR Lexicon</i> | <i>Only in Metadata</i> | |
|--|---------------------------|-------------------------|-----|
| wit | | eichmann | |
| jew | volkswagen | | |
| labor camp | | ig farben | |
| slave labor | | telefunken | aeg |
| minsk ghetto underground | | | |
| | | wallenberg eichmann | |
| bomb birkeneneau | | | |
| sonderkommando auschwitz | | | |
| liber buchenwald dachau | | | |
| jewish kapo | | | |
| kindertransport | | | |
| ghetto life | | | |
| fort ontario refugee camp | | | |
| jewish partisan poland | | | |
| jew shanghai | | | |
| bulgaria save jew | | | |



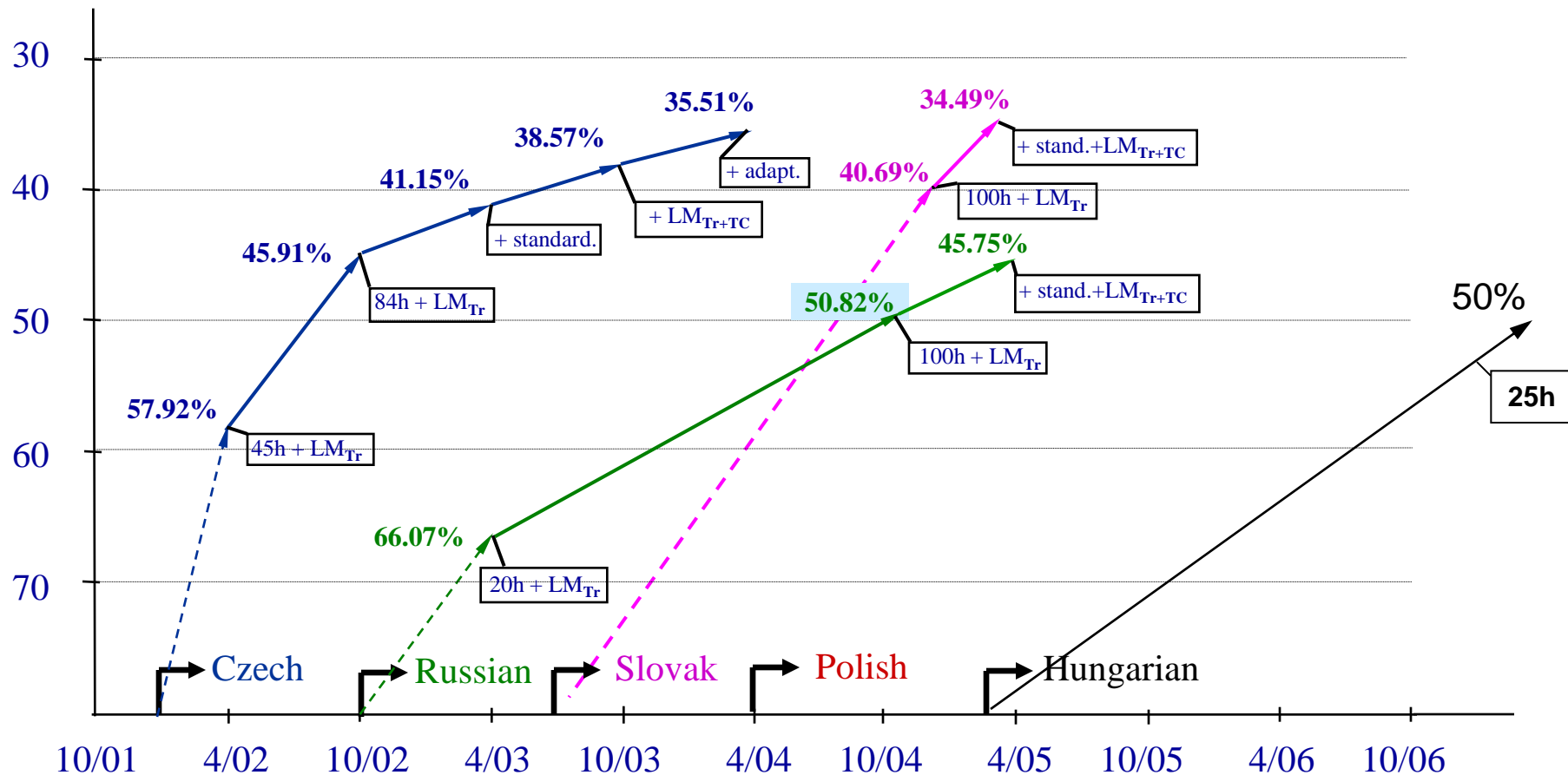
Title queries, adjudicated judgments, Inquiry

Correcting Relevant Segments



Other Languages

WER [%]



2006: Czech Test Collection

- Two sets of speech recognition transcripts
 - Each with both channels recognized
- Queries in Czech or (for CLIR) English
- Unknown-boundary evaluation measure
 - Based only on passage start times
 - Start and end time assessments available

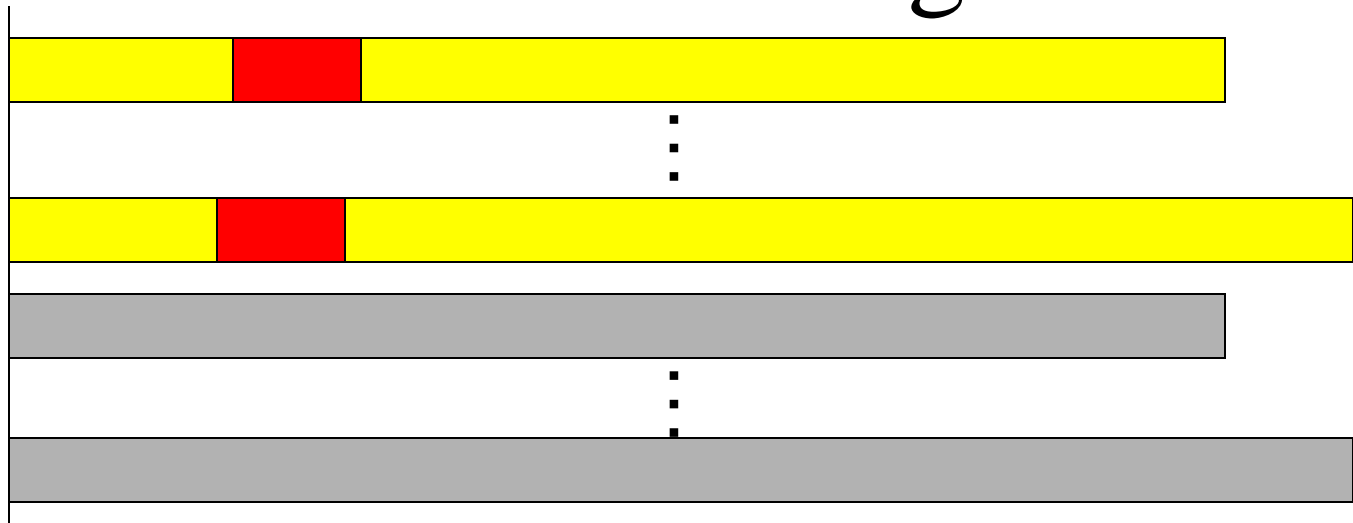
Standard Resources for Czech

- Brain-dead automatic segmentation
 - 400-word passages, with 67% overlap
 - Start time used as DOCNO
- Prague Dependency Treebank (LDC)
 - Czech/English translation lexicon
 - Includes precompiled morphology and probabilities
 - Czech morphology
 - For both formal and colloquial
- Partially translated thesaurus
 - Most commonly used terms, single translation

Interview Usage

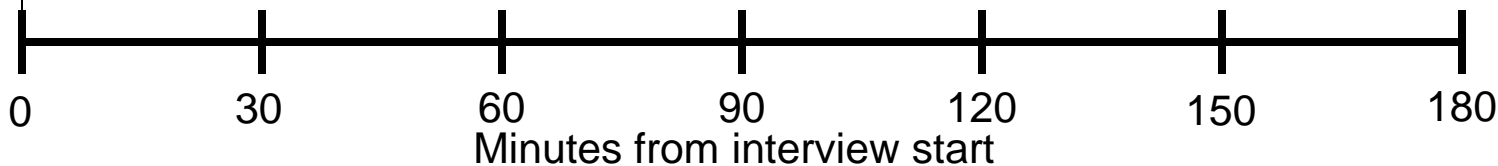
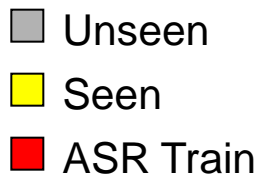
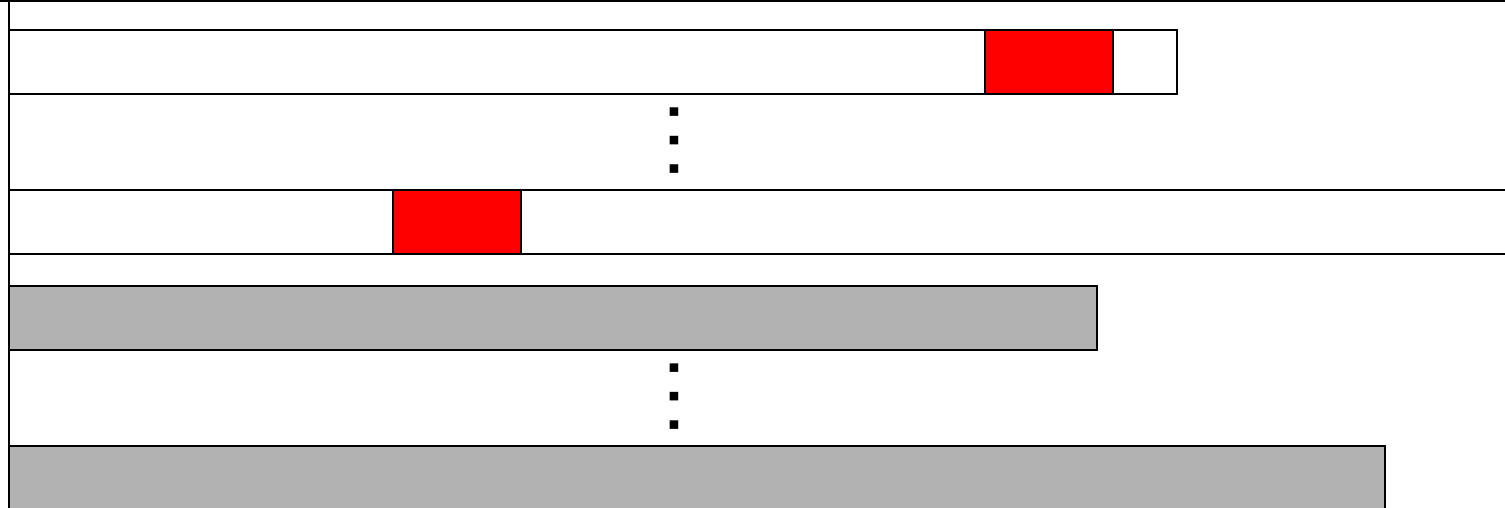
Czech

335 Seen
22 Unseen



English

800 ASR Train
297 IR Eval



Czech Relevance Assessment

- Czech subject experts assessors
 - New assessment system for unsegmented audio
 - 2 assessor training topics available for debugging
- Search-guided assessment run in advance
 - Search based on onset marks + ASR results
 - Summaries from translated thesaurus + ASR
- Highly-ranked after run submission failed
 - Time-base misalignment made this less useful

Czech Assessment System

Relevance Assessment: 1286 - Music in the Holocaust

File Edit Help

Search Assess

Find: V A Find all

Find: V A Find all

music (performance)

Theresienstadt (Czechoslovakia : Ghetto)
cultural and social activities in the ghettos
Czechoslovakia 1942
Czechoslovakia 1943

photographs of interviewee (stills)
photographs (stills) 1940s
photographs (stills) wartime

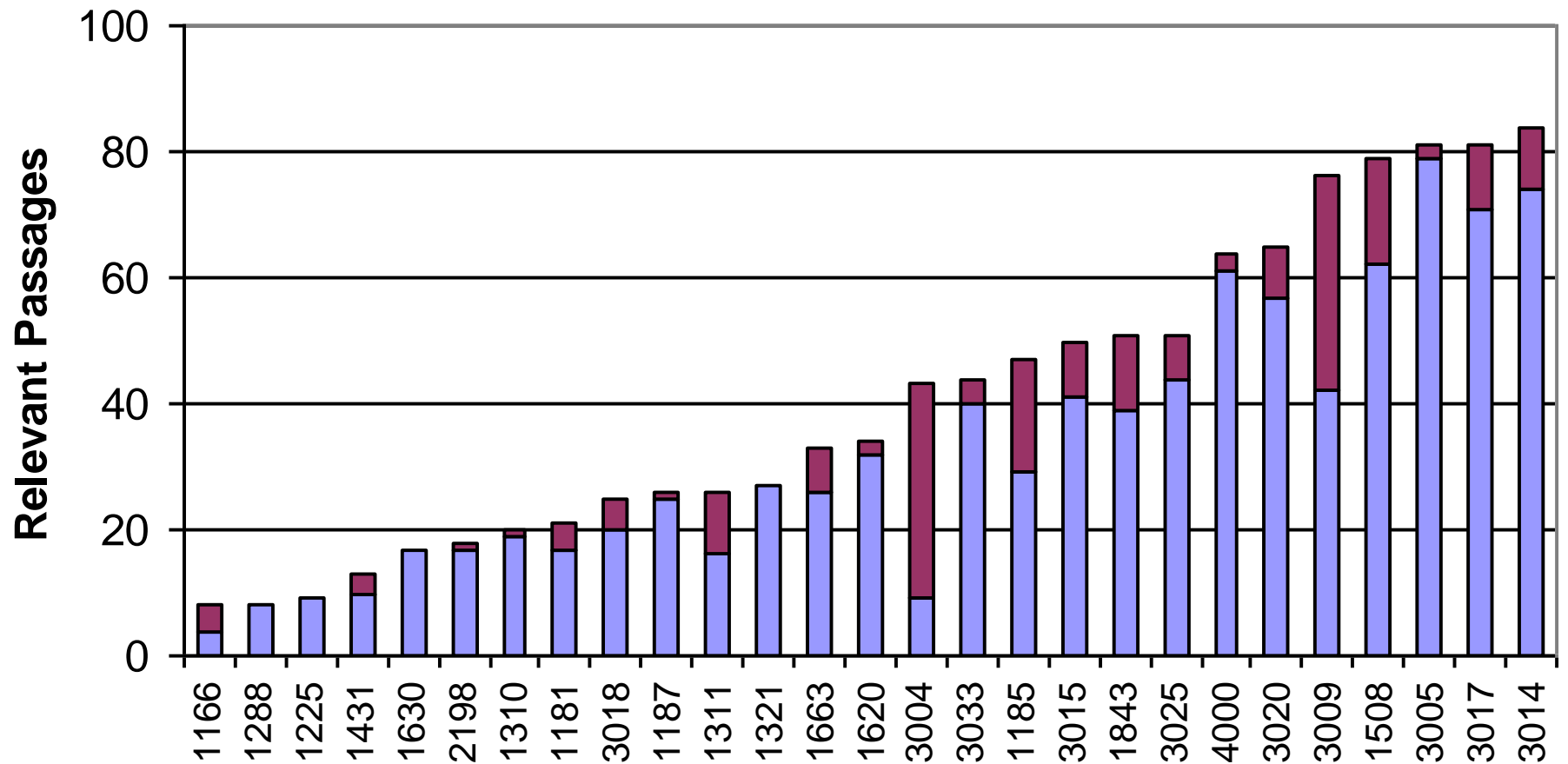
Show: ☒ Locations ☒ Concepts ☒ People

se dozvěděli no ... jistě ano aby už nebyla nikdy válka aby nikdo nezkusil tolik co my jsme zkusili ... no můj manžel tedy byl
hrozně hodný on byl šikovný
on byl starší vo deset let ale uměl vařit a šít a vše pomáhal mi s dětmi mi vychovávat byl strašně hodný ... no ... také si přišel
na svý protože devětatřicátým roce ujel
I do toho izraele to jeli třicetým roku jeli na moři to bylo v devětatřicátým roku to bylo úplně moře zamrzlý a všechno oni tam
zůstali v rumunsku s tou lodí stát tak potom ty ru- rumunský židi se jako složili
a dali jim lepší lod' no a oni potom přijel do izraele ale také málo lodí co se tam dostalo oni také tam zůstali na moři a on měl
štěstí že tam dojel tak pak byl zavřený angličané
ho zavřeli no a potom asi ho pustili za dva měsíce a on zase byl takový šikovný že i na lodi si přivydělával aby měl lepší jídlo ta
dělala hojky šil z prostěradla víte no tak si
pomáhal ... takže se to přežila ... no sem vdova už to bude teď šest roků no je mi smutno no co můžu
I dělat ... ty se nezmínila nebyla ... ráda bych se tam podívala vždycky bojím když jsou tam ty nepokoje a to ... my máme kdy
jsme tam chtěli je to jsme neměli peníze teď sem
dostala pár korun to odškodnění a zase si netroufám no nevím jestli s těma děvčaty že bysme tam jeli ... se podívat ... ještě
nosí ... no nevím ...
to je můj sen ... emilia to přes tu střešovicovou kancelář asi asi ... organizovalo mahler nebo ne ne to já nevím já
ale neumím také ne ne až nás půjdou milióny všichni proti větru já nevím já to neumím to už je jedenapadesát let ...
I já já to řeknu až nás půjdou milióny všichni proti větru já už nevím já já sem nervózní víte nervy mě ...
na nás tam bylo moc to jsme zpívali a ty esesačky zpívaly s námi oni nevěděli co to je a zpívali s námi oni to byla jedna
esesačka řekla že mám umýt záchod a já sem říkal že to neumím tak
to měla za mě řekli že by tam byli známí herci přece v terezíně tam ty míšený manželství měli čtyřicet dvě sestry ... no ale měla
spoustu věcí vy jste praze ... ano
já sem byla ze zavazadlo žížkov třeba děcko která jako spali jsme vedle sebe a ona měla maminku sebou a ta její maminka to
přežila vrátila se ona pochází z plzně zuzana družičku ona ona byla mladší než
I já no dva roky vzala si toho chleba do jeslí teď ona třicet amu a do ciziny jezdí ... no tak tomu sem byla také na v terezíně
sem znala hodně lidí jako
tam toho jak zpíval národním divadle toho kecala bek byl nebo ten ten ta tepna to sem také znala osobně no no a pak sem
viděla toho brundibára také no vždyť já
sem mě bylo de facto šest roků byli jsme bydleli jsme na budyni víte neznali jsme písničky divadlo nic a to vše v terezíně jsme
jako viděli ... no no tam bylo hodně ... kopecký
také je míšenec ten také tam měl přijít hermanová lžuba to víte ta babička její byla židovka a teď ten tóper jestli díváte se na to
život název tak ten je také žid ... no
I ano hodně jako a říkali dokonce hrušinský že prý měl manželku také židovku ... no a kdo koho také sem tam viděla jak je ta
hlasatelka ježíš jak se jmenuje taková šikovná v televizi
ona umí asi pět nebo čtyři sta má marta skarlantová tak ta je také položidovka s tou sem mluvila v terezíně otec její byl také žid
a oni mají teď divadlo spolu myslím v praze a má bratra ještě
... hajna no jo já nevím koho ještě ... tak nám ... fotku nějakou no ...
dvě sestry ... jeho no dobře ...
I to sem ráda aspoň to ukážu tem svým vnukům že jo a tem dětem oni už nemají zájem akorát nejmladší vnuk ten má zájem vo
něm ...
nějakou rýmu ... tak tato fotka je dělaná v jedenačtyřicátým
roce v roudnici to než sem šla do terezína takhle sem vypadala ... tato fotka je dělaná ještě v jedenačtyřicátým
roce než sem šla do terezína když sem měla buřdu a nosla má ji moji přítelkyně která to fotila a potom byla v mojích ...

00:00:29 Set: Channel 2

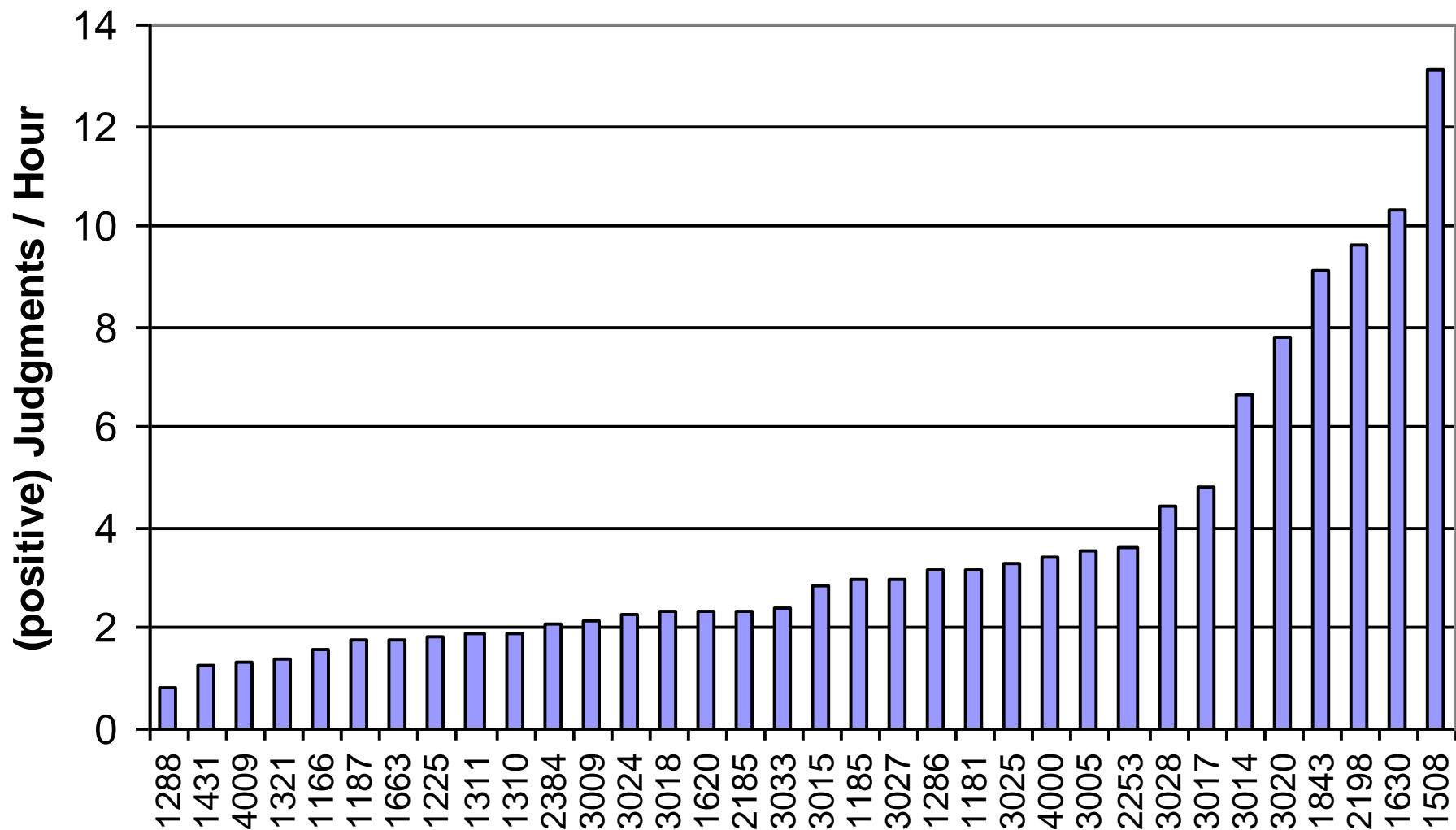
Czech Relevant Passages

Search Guided Highly Ranked



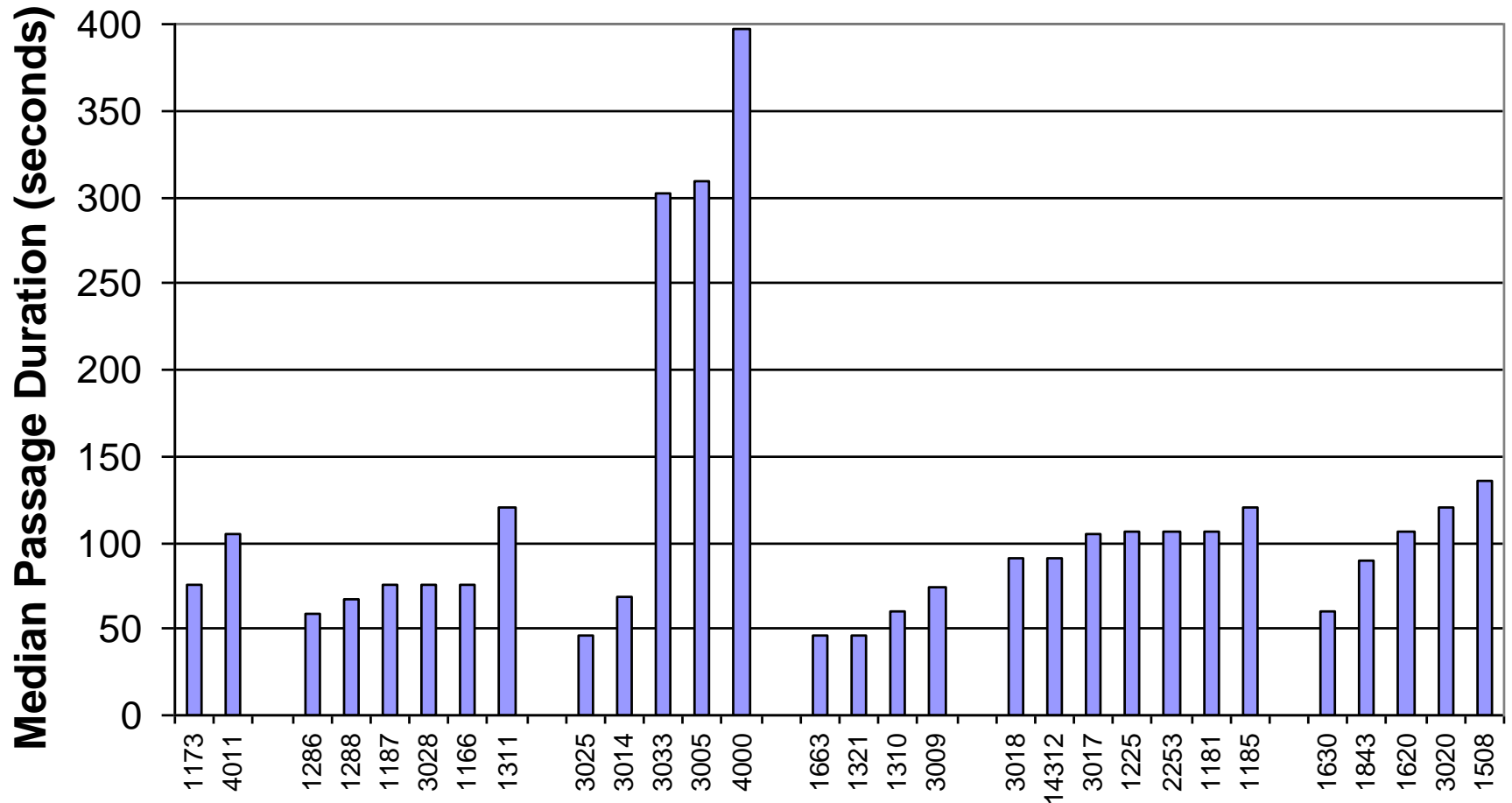
Breakdown not available for topics 1286 (70 passages), 2253 (124 passages)

Czech Judgment Rate

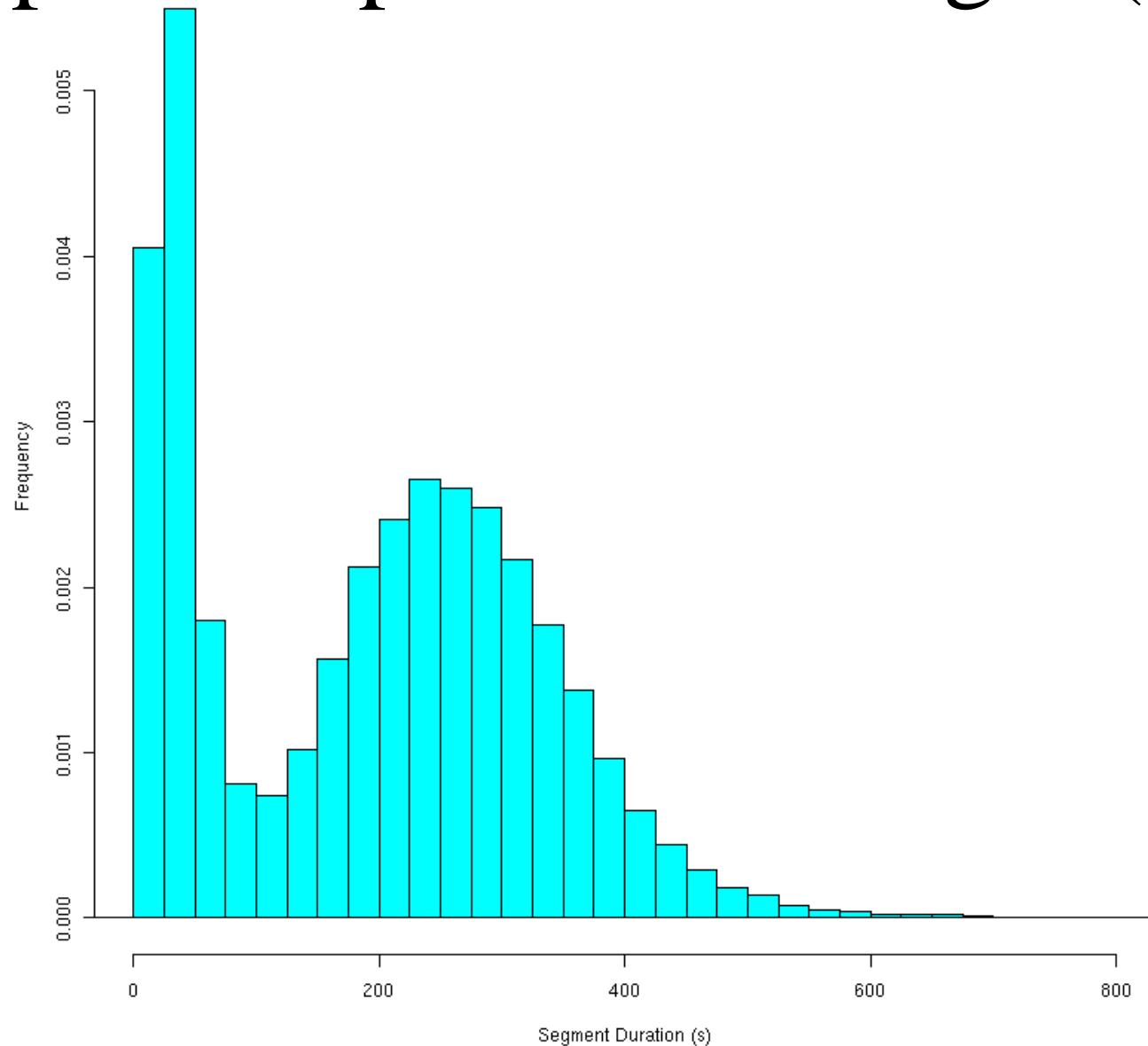


Data not available for topic 3004

Topic-Specific Passages (CZ)

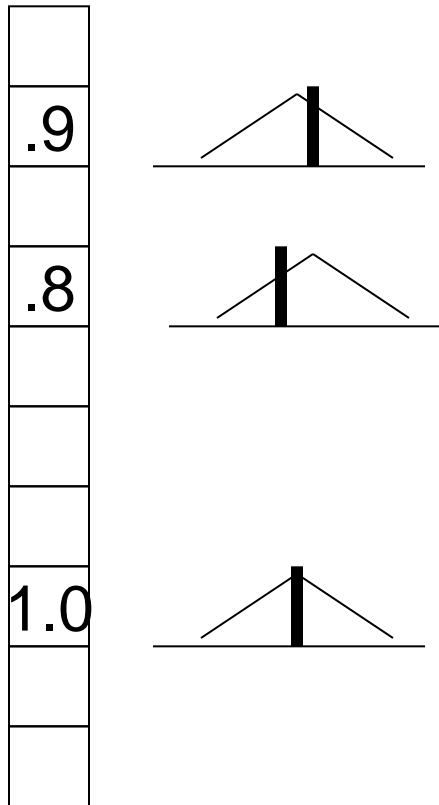


Topic-Independent Passages (EN)

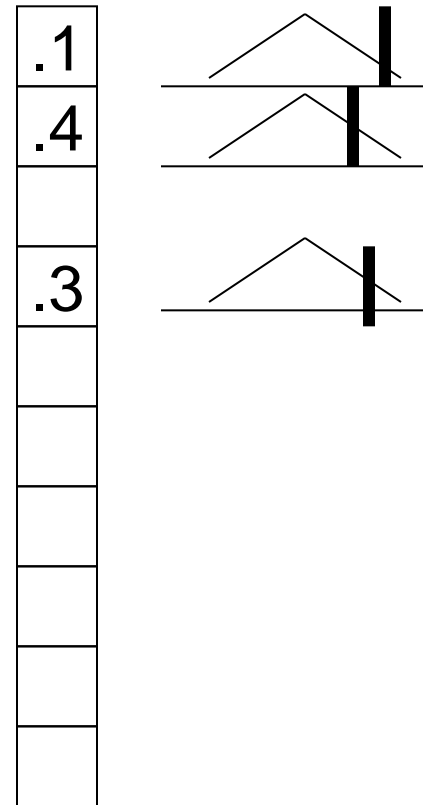


Generalized Average Precision

Key Idea: Weight contribution to average precision by $f(\text{absolute error})$

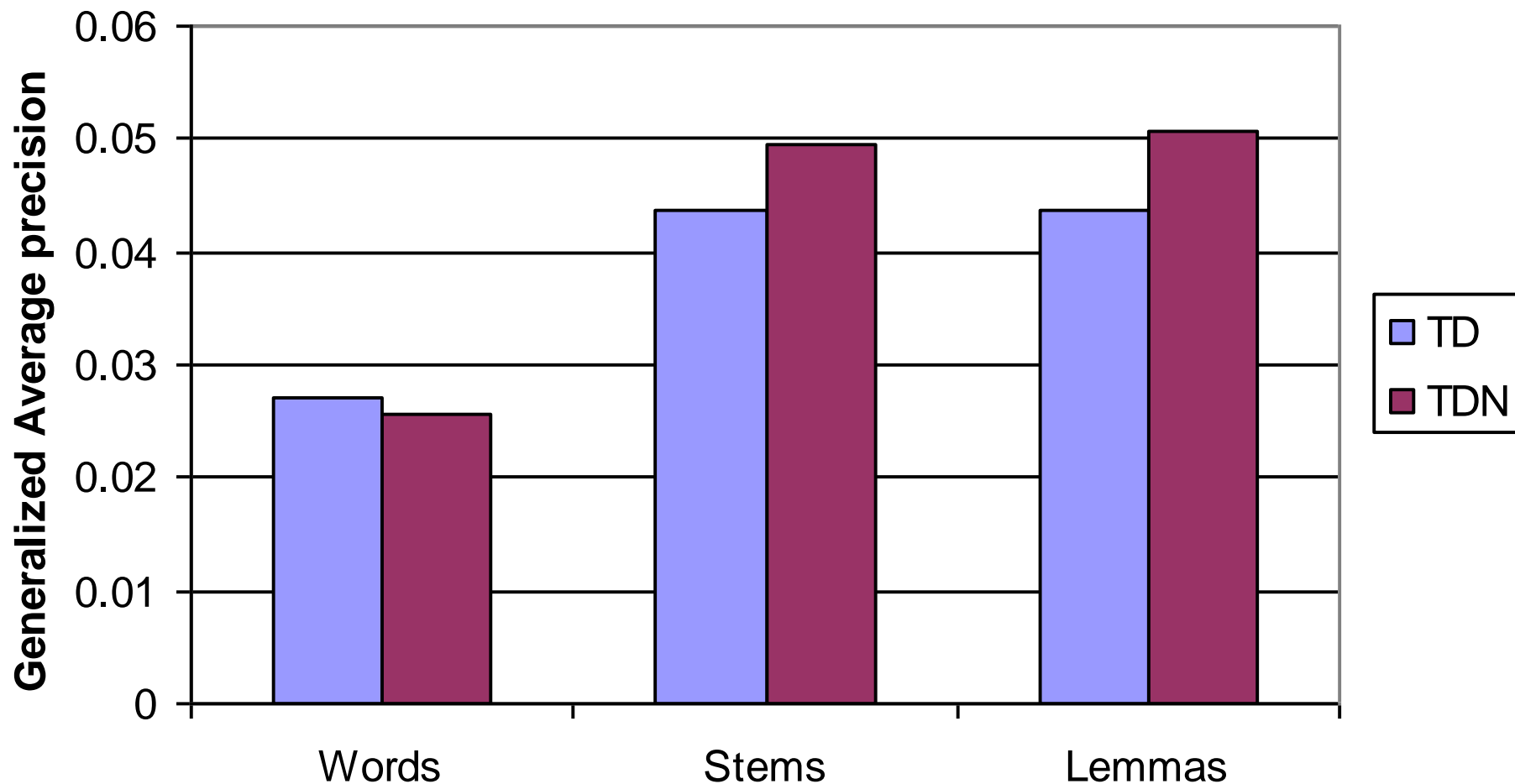


$$\text{GAP} = 0.9/2 + (0.9+0.8)/4 + (0.9+0.8+1.0)/3 = 0.46$$



$$\text{GAP} = 0.1/1 + (0.1+0.4)/2 + (0.1+0.4+0.3)/3 = 0.46$$

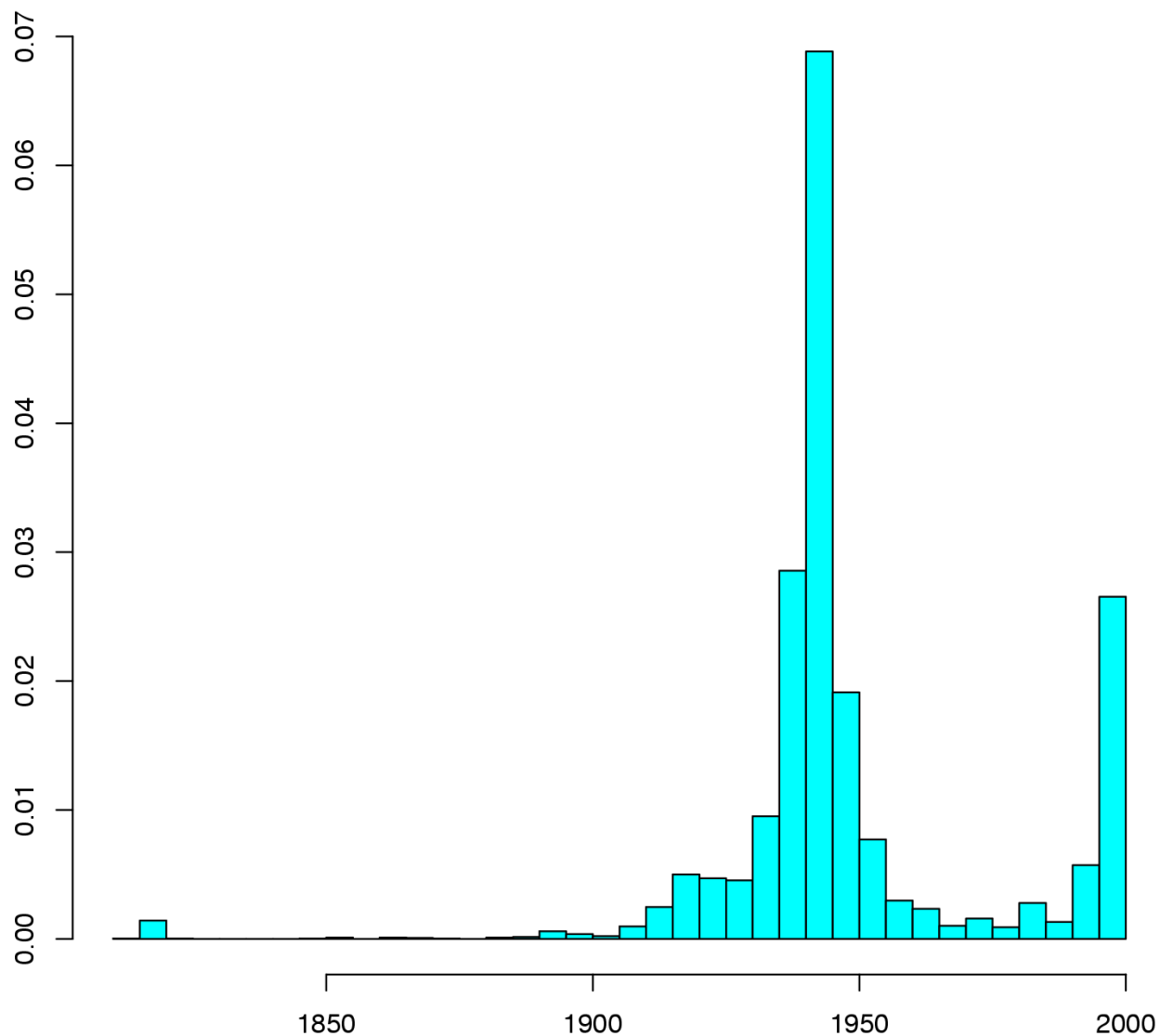
UWB Czech ASR-Only Results



CLEF CL-SR Legacy

- Test collection available from ELDA
 - IR
 - CLIR
 - Topic classification
 - Topic segmentation
- Baseline results
 - Evaluation measures
 - System descriptions

Density of Year Mentions in 35% WER ASR



year

Mean: 2 mentions in every 3 segments

English Information Extraction

| Entity (12) | Event (8) | Relation (34) | | |
|--|---|--|--|--|
| | | Social Rels (12) | Event Args (8) | Bio Facts (14) |
| AGE COUNTRY DATE DATEREFF DURATION GHETTOORCAMP OCCUPATION ORGANIZATION OTHERLOC PEOPLE PERSON SALUTATION | CUSTODY DEATH HIDING LIBERATION MARRIAGE MIGRATION SURVIVAL VIOLENCE | aidgiverOf auntOf cousinOf fatherOf friendOf grandparentOf motherOf otherRelativeOf parentOf siblingOf spouseOf uncleOf | affectedBy agentOf participantIn timeOf travelArranger travelFrom travelPerson travelTo | bornAt bornOn citizenOf diedAt diedOn employeeOf hasProperty locatedAt managerOf memberOf near partOf partOfMany resideIn |

English Information Extraction

| | Train | Dev | Test |
|-----------|-------|-----|------|
| Words | 198k | 73k | 255k |
| Mentions | 43k | 16k | 56k |
| Relations | 7K | 3k | 8k |

- Entity detection
 - In hand-built transcripts: $F = 66\%$
 - In speech recognition: $F = 11\%$
- Relationship detection
 - In hand-built transcripts: $F = 28\%$
 - In speech recognition: $F = 0\%$ (!)

Extractive Summarization

SEGMENT ASR (35% WER):

oh i ll you know are yeah yeah yeah yeah yeah yeah the very why don t we start with you saying anything in your about **grandparents** great **grandparents** well as a small child i remember only one of my grandfathers and his wife his second wife he was selling flour and the type of business it was he didn t even have a store he just a few sacks of different flour and the entrance of an apartment building and people would pass by everyday and buy a chela but two killers of flour we have to remember related times were there was no already baked bread so people had to baked her own bread all the time for some strange reason i do remember fresh rolls where everyone would buy every day but not the bread so that was the business that s how he made a living where was this was the name of the town it wasn t shammay dish he ours is we be and why i as i know in southern poland and alisa are close to her patient mountains it was rather mid sized town and uhhuh i was and the only child and the family i had a governess who was with me all their long from the time i got up until i went to sleep she washed me practice piano she took me to ballet lessons she took me skiing and skating wherever there was else that I was doing being non reach higher out i needed other children to players and the governors were always follow me and stay with me while ours twang that i was a rotten spoiled care from work to do family the youngest and the large large family and everyone was door in the army

KEYWORDS:

Grandfathers, socioeconomic status, cultural and social activities

Przemysl (Poland), Poland 1926 (May 12) - 1935 (May 12), Poland 1935 (May 13) - 1939 (August 31)

QUERY: grandparents

SUMMARY:

why don t we start with you saying anything in your about **grandparents** great **grandparents**...well as a small child i remember only one of my grandfathers and his wife...didn t even have a store he just had a few sacks of different flour and...chela but two killers of flour we have to remember related times were there was...

Outline

- The MALACH Project
 - Learning from our users
 - Evaluating new access technologies
- What's left to do?

What We've Learned

- ASR-based search works
 - Breakeven for hand transcription: ~1,000 hours
 - Breakeven for thesaurus tagging: ~5,000 hours
 - Inaccessible content clusters by speaker
- Supervised classification works
 - Global+local sequence provide useful evidence
- Entity extraction is weak, relationships fail
- Segmentation should be topic-specific

For More Information

- CLEF Cross-Language Speech Retrieval track
 - <http://clef-clsr.umiacs.umd.edu/>
- The MALACH project
 - <http://malach.umiacs.umd.edu/>
- NSF/DELOS Spoken Word Access Group
 - <http://www.dcs.shef.ac.uk/spandh/projects/swag>

Other Possibly Useful Features

- Channel characteristics
 - Cell phone, landline, studio mike, ...
- Accent
 - Another way of grouping speakers
- Prosody
 - Detecting emphasis could help search or browsing
- Non-speech audio
 - Background sounds, audio cues

Human History

Oral Tradition

Writing

Human Future

Writing and Speech