

Starting Up As a Computational Linguist

Anke Holler
University of Heidelberg

Computational Linguistics Fall School
Bochum, 29.10.2005

What do you do next...

... after graduation?



© Anke Holler 2005

2

Agenda

1. Job Market & Key Qualifications
2. Trends in CL & Language Technology
3. Start-Up or Large Company?
4. Conclusions



© Anke Holler 2005

3

Agenda

1. **Job Market & Key Qualifications**
2. Trends in CL & Language Technology
3. Start-Up or Large Company?
4. Conclusions



© Anke Holler 2005

4

Bad news

- Current job market is highly competitive.
- In Germany, most available jobs seem to be academic ones.



Good news

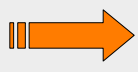
- There is a niche for a non-academic career: **speech and language technology**
- International opportunities



- Interdisciplinary qualifications

Key Qualifications

- Depth in linguistics
- Programming and scripting skills
 - Java/ C++ , Perl, Prolog, XML, JavaScript
- Knowledge of hot applications
- Soft skills
 - good communication skills, capacity for teamwork, ability to work independently, abstract thinking, ability to work in a structured way, critical faculty, etc.
- On-the-job training and practical experiences

 Essential to compete on the job market with Computer Science graduates

Jobs by qualification type

- Programming skills alone are not enough.
- There are non-academic jobs at the universities.



What fits best?

What do I like most in CL?

What do I hate most in CL?

What are my strenghts?

What are my weaknesses?

Which professional objectives do I have?



Possible career entries

■ Lingware developer

- modeling linguistic data
- creating language ressources

■ Software developer

- software engineering
- developing NLP-based systems

■ Salesperson

- presenting and selling LT
- explaining basic technology
- tool training

■ Consultant

- agent for language technology
- providing applicative solutions

A typical task

■ Software developer

Design and implement a fast and robust transducer that can be easily integrated into an existing software architecture.

■ Lingware developer

Write a regular grammar to extract financial information (sales figures, profitability, growth), sales information (market shares, number of customers) or stock market information (capitalization, trends) out of a set of documents.

■ Consultant

Conceptualize a knowledge management system for a large company and manage a respective customer project.

■ Salesperson

Get and keep clients for the company. Distribute products.

Possible career entries

There is a lot in between:

- (Technical) Support
- Presales
- Product Management
- Training
- Document Processing/Technical Writing/Translation
- ...

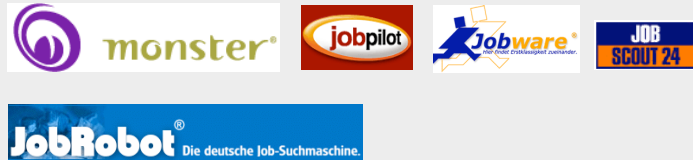


Watch out for job descriptions mentioning these areas as well!

Where to search active job ads?

- Job area of the **linguist list**:
<http://www.linguistlist.org/jobs/index.html>

- Several **job markets and job search engines**



- Web pages of **potential employers**

How to contact a company?

- To come in touch:
 - Internships
 - Graduate recruiting fairs
 - Bonding contact fairs
 - Post your profile to a job market
- To apply for a job:
 - React to an active job announcement
 - Use online application forms
 - Apply blindly
 - Talk to people

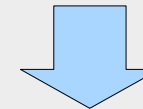
Agenda

1. Job Market & Key Qualifications
2. Trends in CL & Language Technology
3. Start-Up or Large Company?
4. Conclusions



Emergence of the language engineering paradigm

- Demand for language industry products
 - to assist the traditional linguistic professions (translation, language teaching, etc.)
 - to develop new language processing applications (natural language interfaces, speech input and output, document retrieval and indexing, etc.)



Requires development of robust language processing components, capable of dealing with real texts in concrete information and communication systems

= language technology

Language technology (LT) is ...

... the cover term for all information technologies specialized for dealing with text and speech in human language. It is also the field of engineering in which LT methods and applications are developed.

■ **Getting an Insight:** <http://www.lt-cc.org/index-e.html>

German Competence Center for Speech and LT:

- Virtual Information Center LT-World
- German Demonstration Center for LT Systems
- Evaluation Center for LT applications



Stakeholders of LT

- Computer programmers, NLP experts and computational linguists
 - conduct basic research
 - create language engineering solutions
- Users of language engineering solutions
 - professional application experts: translators, localizers, terminologists, and interpreters
 - non-specialist users: bankers, layers, health professionals, etc.

Recent Trends in CL & LT



Ruslan Mitkov (ed.) (2003).
The Oxford Handbook of Computational Linguistics.

Now available as paperback!

Survey of the State of the Art in Human Language Technology.
Editorial Board. (1997)

http://www.ltworld.org/HLT_Survey/master.pdf

2nd. edition currently in preparation



Principal Application Areas

Information Extraction / Information Retrieval
Authoring Tools
Language Analysis / Language Understanding
Knowledge Representation and Discovery
Spoken Language Input / Written Language Input
Natural Language Generation
Spoken Language Output Technologies
Discourse and Dialogue
Multilinguality
Multimodality
Coding and Compression
Language Resources
Evaluation

Topics at ACL 2005

■ Main Conference Sessions:

Corpus Annotation

Generation

Information Extraction

Lexical Acquisition from Corpora

Machine Learning and Statistical Methods

Machine Translation

Segmentation, Tagging, and Semantic Role Labeling

Semantics

Speech and Language Modeling

Summarization

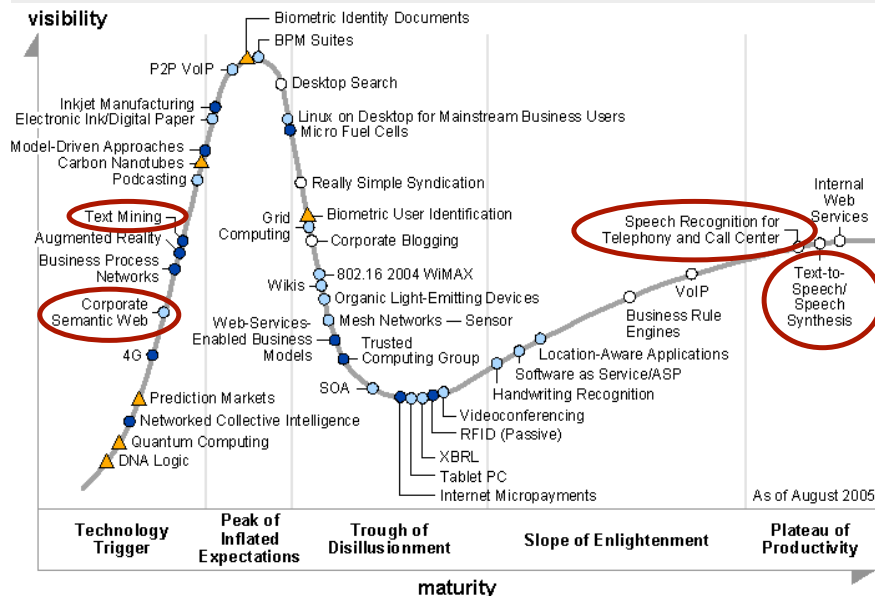
Parsing

Word Sense Disambiguation

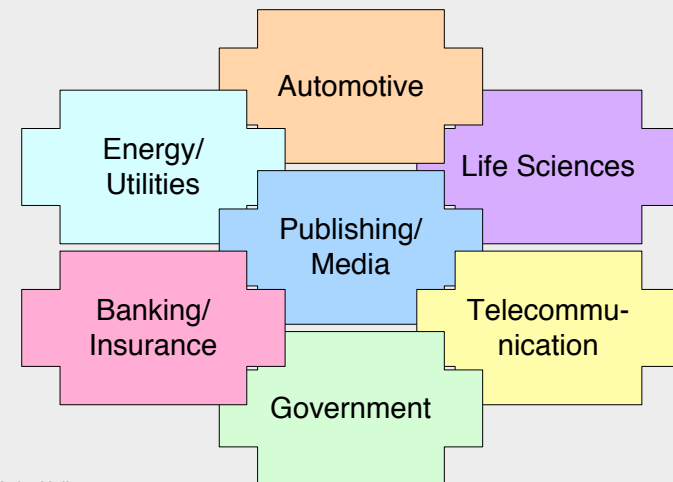
Internet Survey of GLDV

What is the big thing in NLP?

- Speech Understanding 32.26 %
- Text/Data Mining 19.35 %
- Ontologies 19.35 %
- Neurolinguistics 16.13 %



Industries applying LT



Challenges to the CL community

- Combination of new empirical corpus-based methods with traditional symbolic methods
- From purely rule-based or purely statistical methods to hybrid methods
- Creation of technically mature LT for hot application areas:
 - multilingual processing, learning environments, multimodal communication, bioNLP, spam filtering, security, etc.

Challenges for LT companies

- Development of commercial products that
 - cover frequent linguistic phenomena
 - are robust, fast and user-friendly
 - reduce global costs and time-to market
 - increase the return on investment (ROI)

Agenda

1. Job Market & Key Qualifications
2. Trends in CL & Language Technology
3. Start-Up or Large Company?
4. Conclusions



Potential Employers



My experiences



- IBM Development Laboratory, Böblingen
 - IBM is the world's largest information technology company
 - Number of employees worldwide: 319,273
 - Lab in BB engaged in research and development of hardware and software technologies
 - Lab in BB employs more than 1700
- TEMIS GmbH, Heidelberg
 - a software company designing, developing and distributing corporate Text Mining solutions
 - founded in September 2000 by a team of managers, researchers and consultants from IBM
 - employs 50 people and operated subsidiaries in France, Germany, Italy, UK and USA.



Chances and Risks

- Start-Up
 - Company
 - One innovative technology
 - Flat structure
 - Dependent on Investors
 - Staff
 - Variety of tasks
 - Training on-the-job
 - Less opportunities for advancement within the company
 - Sometimes better salaries
 - Compensation mostly includes stock options
- Large Company
 - Company
 - Several business units
 - Hierarchically structured
 - Relevance of LT depends on general strategy
 - Staff
 - Specific task
 - Trainee programs/ Professional trainings
 - Job rotation within the company possible
 - More social security, e.g. pension plan

You need a Start-Up mentality

- Willingness to be extremely hands-on
- Being proactive and optimistic
- Being flexible (in time and location)
- Willingness to work on a variety of tasks
- Ability to develop projects to extremely tight timescales
- Ability to work under pressure
- Being independent of securities
- Willingness to take on responsibility
- Being entrepreneurial

Ask yourself ...

... whether you prefer to work

on an ocean liner



or on a sail boat !

Agenda

1. Job Market & Key Qualifications
2. Trends in CL & Language Technology
3. Start-Up or Large Company?
4. **Conclusions**



Conclusions

- CL enables innovative products in various sectors of industry.
- LT has a huge potential in future markets.



- Many careers are open to graduates of Computational Linguistics.
- Start-ups, large-scale enterprises and research institutes offer manifold non-academic job opportunities.
- Identify your personal preferences: What suits best to you?
- Take the decision into your own hands!

Focus Areas of Language Engineering

- Interface and support tools
- Shared language resources
 - include text databases and corpora, translation memory corpora, speech corpora, reference and encyclopedic products, lexicographical and terminological databases, and field-specific thesauri and classification systems
- Standards
 - character sets like UNICODE
 - various text markup, formatting standards, and file conversion standards

- Speech Communication
- Adaptive technology
- Education
- Document Generation and Management
- Translation
- Ontological systems

Principal application Areas

- **Information Extraction**[HLT-Survey Section Answer Extraction](#)[Information Extraction](#)[Multimedia Information Extraction](#)[Named Entity Recognition](#)[Relation Extraction](#)[Summarisation](#)[Text Data Mining](#) **Information Retrieval**[HLT-Survey Section Categorization](#)[Clustering](#)[Information Retrieval](#)[Multilingual Information Retrieval](#)[Multimedia Retrieval](#)[Presentation and Visualisation](#)[Relevance Ranking](#)[Speech Retrieval](#)[Topic Detection](#) **Authoring Tools**[HLT-Survey Section Automatic Hyperlinking](#)[Language Checking](#)[Spell Checking](#)[Structure-Based Authoring Assistants](#) **Language Analysis**[HLT-Survey Chapter](#)[Categorial Grammar](#)[Dependency Grammar](#)[Government and Binding Theory / Minimalist Framework](#)[Grammar Models and Formalisms](#)[Head-driven Phrase Structure Grammar](#)[Lexical-Functional Grammar](#)[Lexicons for Constraint-Based Grammars](#)[Morphological Analysis](#)[Natural Language Parsing](#)[Optimality Theory in Syntax](#)[Parsing Techniques](#)[Part-of-speech Tagging](#)[Probabilistic Context-free Grammars](#)[Shallow Parsing](#)[Systemic Functional Linguistics](#)[Tokenization and Segmentation](#)[Tree Adjoining Grammar](#) **Language Understanding**[HLT-Survey Chapter](#)[Computational Pragmatics](#)[Computational Psycholinguistics](#)[Computational Semantics](#)[Word Sense Disambiguation](#) **Knowledge Representation and Discovery**[HLT-Survey Chapter](#)[Automatic Hyperlinking](#)[Knowledge Discovery](#)[Ontologies](#)[Semantic Web](#) **Spoken Language Input**[HLT-Survey Chapter](#)[Acoustic Modelling in Speech Recognition](#)[Emotion Recognition](#)[Language Modelling](#)[Prosody Information Processing](#)[Signal Analysis and Representation](#)[Speaker Recognition](#)[Speech Recognition](#)[Spoken Language Understanding](#) **Written Language Input**[HLT-Survey Chapter](#)[Document Image](#)

© Anke Holler 2005

37

Industry sectors using NLP

- aims to provide an overview of current trends and challenges in putting Text Mining approaches to work in a variety of scenarios. More and more these methods are used large scale analysis of patent documents
 - pharmacy, chemistry and biotechnology
- Automatic processing of multilingual medical terminology:

© Anke Holler 2005

38

Statistical NLP

- Three areas of recent progress:
 - part-of-speech tagging
 - stochastic parsing
 - lexical semantics

© Anke Holler 2005

39

Main application areas

- Question Answering
- Automatic Indexing
- Text Data Mining
- Text Summarization
- Natural Language Interaction
- Natural Language in Multimodal & Multimedia Systems
- NLP in Computer-Assisted Language Learning
- Speech Processing

© Anke Holler 2005

40

Agenda

1. Job Market & Key Qualifications
2. Trends in CL & Language Technology
3. Start-Up or Large Company?
4. **Some Practical Hints**
5. Conclusions

