

FLAG: Flexible language and grammar checking

DFKI, Saarbrücken, Germany

contact: Andrew.Bredenkamp@dfki.de



Overview

- ❑ The Flag Project
- ❑ Our goals
- ❑ Overview of functionality
- ❑ Implementation – how it works
- ❑ Integration – how it can be used



The Flag Project

- ❑ Flexible Language and Grammar checking
- ❑ 3-year BMBF research project
- ❑ Main aim:
to build up competence in the domain of
“controlled language”, to meet the
requirements of different clients.



Controlled Languages

- ❑ Offering clear, easily understandable communication
- ❑ Helping to establish corporate identity
- ❑ Working with a restricted controlled and consistent vocabulary
- ❑ Problem: how can this best be guaranteed?...
- ❑ ... Automatically



Our goals

- ❑ To produce a platform for the rapid development of user-specific language checking applications
- ❑ Why?
 - Language checking applications:
 - Are always client- or even user-specific
 - Must be integrated into existing workflows (document production environments)
 - Should use “appropriate technology”



Functionality

- ❑ Terminology extraction methods
 - (User) adaptability
 - Text and domain specific heuristics
 - Automatic hyperlinking with translations and definitions



Functionality II

- ❑ Terminology checking
 - Hierarchical lexica
 - Lemma-based lookup
 - Checking for “illegal variants” *
 - Domain-specific ontology (thesaurus) access*
 - Automatic hyperlinking

*work in progress



Functionality III

- ❑ Integration of a wide range of language processing components
- ❑ Language checking
 - Based on the notion of “triggers”
 - Identification of candidates (as cheaply as possible)
 - Confirmation of hypotheses by focused deeper processing
- ❑ Formalism for error description
- ❑ Specific rules for specific (error) phenomena



Error description

- ❑ As “lean” a formalism as possible
- ❑ As “declarative” as possible
- ❑ Pattern-matching over linguistic descriptions
- ❑ Constraints over linguistic descriptions
- ❑ Interpreted rule sets
- ❑ Full debugging environment with tracing



Error example

- ❑ #ERROR reject_passives
- ❑ #OBS
- ❑ @werden ::= [MORPH.LEMMA
" ^werden\$ "];
- ❑ @participle ::= [POS "VVPP"];
- ❑ @komma ::= [TOK ","];
- ❑ @comp ::= [POS "KOUS"];

- ❑ #RULES
- ❑ /* A form of "werden" together with a past participle is very likely a passive when appearing in the same (sub)clause, as indicated by the topology parser. */
- ❑ TRIGGER(80) ==
@werden^1 []* @participle^2
&& sin(^1, "SATZ"^3)
&& sin(^2, "SATZ"^3)
-> \$werden^1, \$participle^2;
- ❑ /* This trigger acts as a fallback point in case the parser could not assign a sentence structure. */
- ❑ TRIGGER(70) ==
@werden^1 []* @participle^2
-> \$werden^1, \$participle^2;
- ❑ NEG_EV(30) == \$werden []* @komma
@comp []* \$participle;



Implementation

- ❑ Tools for checking:
 - Terminology
 - Controlled language
 - Grammar
- ❑ “appropriate technology”
 - Expensive processes are only used where necessary
 - Checking is :
 - Resource-sensitive
 - Resource independent
 - Language independent



Resources

- ❑ Multilingual (German, English French, ... (Japanese))
- ❑ Tokeniser (not trivial)
- ❑ High-precision Part-of-Speech Tagger
- ❑ Broad-coverage morphological processing components
- ❑ Stemmer
- ❑ Phrase boundary recognition (partial parsing)



Integration

- ❑ How can the system be used in existing workflows?
- ❑ Platform independence:
 - Runs under Windows (NT, 9X) and Unix (Solaris)
 - Main functionality available as libraries (DLLs, shared libraries) for tight integration
 - Accepts all standard formats - SGML/XML, RTF, HTML, text
- ❑ Implementation in Java und C(++)



Built-in linguistic intelligence

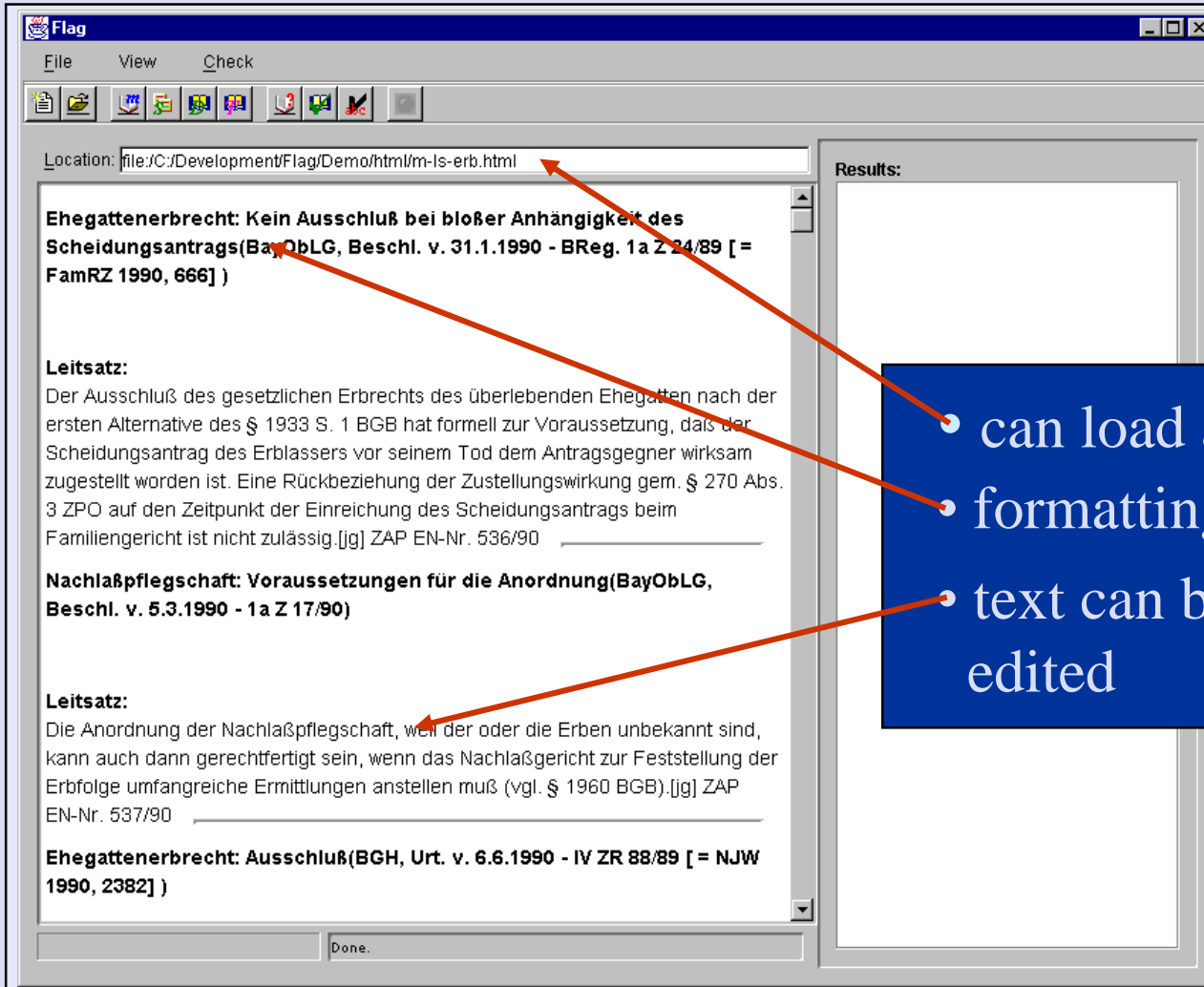
The screenshot shows the FLAG software interface. The main window displays a document with legal text. The text includes sections like "Ehegattenerbrecht: Kein Ausschluß bei bloßer Anhängigkeit des Scheidungsantrags", "Leitsatz:", and "Nachlaßpflegschaft: Voraussetzungen für die Anordnung". The word "Erblassers" is highlighted in the text. An orange arrow points from this word to the "Results:" window on the right. The "Results:" window shows the following XML-like structure:

```
Results:
<TOKEN STRING="Erblassers">
  <READING STEM="Erblasser" CAT="Noun">
    <INFLECTION
      gender="masc"
      number="singular"
      case="gen."
    />
  </READING>
</TOKEN>
```

Linguistic
Analysis:
identification of
terms as lemmas



Checking tools



- can load any URL
- formatting retained
- text can be browsed or edited



Checking tools II

The screenshot shows the FLAG web application interface. The main window displays a document with several paragraphs of text. The text is color-coded: unknown words are in red, general terms are in blue, and domain-specific terms are in pink. The results panel on the right is currently empty. A blue callout box on the right side of the screenshot contains three bullet points: 'Unknown words (red)', 'General terms (blue)', and 'Domain-specific terms (pink)'. Red arrows point from these callouts to the corresponding highlighted text in the document.

Location: file:/C:/Development/Flag/Demo/html/m-ls-erb.html

Ehegattenerbrecht: Kein **Ausschluß** bei bloßer **Anhängigkeit** des **Scheidungsantrags** (**BayObLG, Beschl. v. 31.1.1990 - BReg. 1a Z 24/89 [= FamRZ 1990, 666]**)

Leitsatz:
Der **Ausschluß** des gesetzlichen **Erbrechts** des überlebenden **Ehegatten** nach der ersten Alternative des § 1935 S. 1 BGB hat formell zur Voraussetzung, daß der **Scheidungsantrag** des **Erblässers** vor seinem **Tod** dem **Antragsgegner** wirksam zugestellt worden ist. Eine **Rückbeziehung** der **Zustellungswirkung** gem. § 270 Abs. 3 ZPO auf den **Zeitpunkt** der **Einreichung** des **Scheidungsantrags** beim **Familiengericht** ist nicht zulässig. [jg] ZAP EN-Nr. 536/90

Nachlaßpflegschaft: Voraussetzungen für die **Anordnung** (**BayObLG, Beschl. v. 5.3.1990 - 1a Z 17/90**)

Leitsatz:
Die **Anordnung** der **Nachlaßpflegschaft**, weil der oder die **Erben** unbekannt sind, kann auch dann **gerechtfertigt** sein, wenn das **Nachlaßgericht** zur **Feststellung** der **Erfolge** umfangreiche **Ermittlungen** anstellen muß (vgl. § 1960 BGB). [jg] ZAP EN-Nr. 537/90

Ehegattenerbrecht: **Ausschluß** (BGH, **Urt. v. 6.6.1990 - IV ZR 88/89 [= NJW 1990, 2382]**)

Done.

- Unknown words (red)
- General terms (blue)
- Domain-specific terms (pink)



Automatic hyperlinking

Location: file://C:/Development/Flag/Demo/html/m-ls-erb.html

Ehegattenerbrecht: Kein Ausschluß bei bloßer Anhängigkeit des Scheidungsantrags(BayObLG, Beschl. v. 31.1.1990 - BReg. 1 a Z 24/89 [= FamRZ 1990, 666])

Leitsatz:
Der **Ausschluß** des gesetzlichen **Erbrechts** des überlebenden **Ehegatten** nach der ersten Alternative des § 1933 S. 1 BGB hat formell zur Voraussetzung, daß der **Scheidungsantrag** des **Erblassers** vor seinem **Tod** dem **Antragsgegner** wirksam zugestellt worden ist. Eine **Rückbeziehung** der Zustellungswirkung gem. § 270 Abs. 3 ZPO auf den **Zeitpunkt** der **Einreichung** des **Scheidungsantrags** beim **Familiengericht** ist nicht zulässig. [jg] ZAP EN-Nr. 536/90

Nachlaßpflegschaft: Voraussetzungen für die Anordnung(BayObLG, Beschl. v. 5.3.1990 - 1 a Z 17/90)

Leitsatz:
Die **Anordnung** der **Nachlaßpflegschaft**, weil der oder die **Erben** unbekannt sind, kann auch dann gerechtfertigt sein, wenn das **Nachlaßgericht** zur **Feststellung** der **Erbfolge** umfangreiche **Ermittlungen** anstellen muß (vgl. § 1960 BGB). [jg] ZAP EN-Nr. 537/90

Ehegattenerbrecht: Ausschluß(BGH, Ur. v. 6.6.1990 - IV ZR 88/89 [= NJW 1990, 2382])

Results:

- § 1961 **Nachlaßpflegschaft auf Antrag**
Das Nachlaßgericht hat in den Fällen des § 1960 Abs. 1 einen Nachlaßpfleger zu bestellen, wenn die Bestellung zum Zwecke der gerichtlichen Geltendmachung eines Anspruchs, der sich gegen den Nachlaß richtet, von dem Berechtigten beantragt wird.
- § 1962 **Zuständigkeit des Nachlaßgerichts**
Für die **Nachlaßpflegschaft** tritt das Vormundschaftsgericht des Nachlaßgerichts.
- § 1975 **Nachlaßverwaltung - Nachlaß**
Die Haftung des Erben für die Nachlaßverbindlichkeiten beschränkt sich auf den Nachlaß, wenn eine **Nachlaßverwaltung** zum Zwecke der Befriedigung der Nachlaßgläubiger (Nachlaßverwaltung) angeordnet oder der Nachlaßkonkurs eröffnet ist.
- § 2017 **Nachlaßpflegschaft**
Wird vor der Annahme der Erbschaft zur Verwaltung des Nachlasses ein Nachlaßpfleger bestellt, so beginnen die im § 2014 und im § 2015 Abs. 1 bestimmten Fristen mit der Bestellung.

Definitions:
Terms in context



Automatic hyperlinking II

The screenshot shows the FLAG application window with a menu bar (File, View, Check) and a toolbar. The main window displays a legal document with several terms highlighted in blue and underlined, indicating they are hyperlinks. On the right side, a 'Results:' panel lists translations for these terms in English, Dutch, French, and Italian. Red arrows point from the hyperlinks in the text to their respective entries in the translation list.

Location: file://C:/Development/Flag/Demo/html/m-ls-erb.html

Ehegattenerbrecht: Kein **Ausschluß** bei bloßer **Anhängigkeit** des **Scheidungsantrags** (BayObLG, Beschl. v. 31.1.1990 - BReg. 1 a Z 24/89 [= FamRZ 1990, 666])

Leitsatz:
Der **Ausschluß** des gesetzlichen **Erbrechts** des überlebenden **Ehegatten** nach der ersten Alternative des § 1933 S. 1 BGB hat formell zur Voraussetzung, daß der **Scheidungsantrag** des **Erblassers** vor seinem **Tod** dem **Antragsgegner** wirksam zugestellt worden ist. Eine **Rückbeziehung** der Zustellungswirkung gem. § 270 Abs. 3 ZPO auf den **Zeitpunkt** der **Einreichung** des **Scheidungsantrags** beim **Familiengericht** ist nicht zulässig. [jg] ZAP EN-Nr. 586/90

Nachlaßpflegschaft: Voraussetzungen für die **Anordnung** (BayObLG, Beschl. v. 5.3.1990 - 1a Z 17/90)

Leitsatz:
Die **Anordnung** der **Nachlaßpflegschaft**, weil der oder die **Erben** unbekannt sind, kann auch dann gerechtfertigt sein, wenn das **Nachlaßgericht** zur **Feststellung** der **Erbfolge** umfangreiche **Ermittlungen** anstellen muß (vgl. § 1960 BGB). [jg] ZAP EN-Nr. 537/90

Ehegattenerbrecht: **Ausschluß** (BGH, Urt. v. 6.6.1990 - IV ZR 88/89 [= NJW 1990, 2382])

Results:

ENGLISH:

- testator ; devisor ; bequeather ; the deceased
- deceased person
- restoration of gifts and legacies

DUTCH:

- erflater ; de cujus ; decujus ; de cuius
- erflater; de cujus
- erfgenaam
- inbrenq; collatie

FRENCH:

- testateur ; de cujus ; instituant
- de cujus; de cujus successione; de cujus bonis agitur; défunt
- testateur
- testateur; de cujus
- défunt; de cujus
- hoir
- rapport des dons et des legs
- reprenant

ITALIAN:

- testatore; de cuius here
- defunto; de cuius
- erede in linea diretta
- conferimento di beni alla massa

translations:
automatically
acquired
(non-validated)



Automatic Hyperlinking III

Ehegattenerbrecht: Kein Ausschluß bei bloßer Anhängigkeit des Scheidungsantrags(BayObLG, Beschl. v. 31.1.1990 - BReg. 1a Z 24/89 [= FamRZ 1990, 666])

Leitsatz:
Der Ausschluß des gesetzlichen Erbrechts des überlebenden Ehegatten nach der ersten Alternative des § 1933 S. 1 BGB hat primär zur Voraussetzung, daß der Scheidungsantrag des Erblassers vor seinem Tod dem Antragsgegner wirksam zugestellt worden ist. Eine Rückbeziehung der Zustellungswirkung gem. § 270 Abs. 3 ZPO auf den Zeitpunkt der Einreichung des Scheidungsantrags beim Familiengericht ist nicht zulässig.[jg] ZAP EN-Nr. 536/90

Nachlaßpflegschaft: Voraussetzungen für die Anordnung(BayObLG, Beschl. v. 5.3.1990 - 1a Z 17/90)

Leitsatz:
Die Anordnung der Nachlaßpflegschaft, weil der oder die Erben unbekannt sind, kann auch dann gerechtfertigt sein, wenn das Nachlaßgericht zur Feststellung der Erbfolge umfangreiche Ermittlungen anstellen muß (vgl. § 1960 BGB). [jg] ZAP EN-Nr. 537/90

Ehegattenerbrecht: Ausschluß(BGH, Urt. v. 6.6.1990 - IV ZR 88/89 [= NJW 1990, 2382])

WDR

EURODICAUTOM

Search query: Erblasser LOOK UP

HitList:

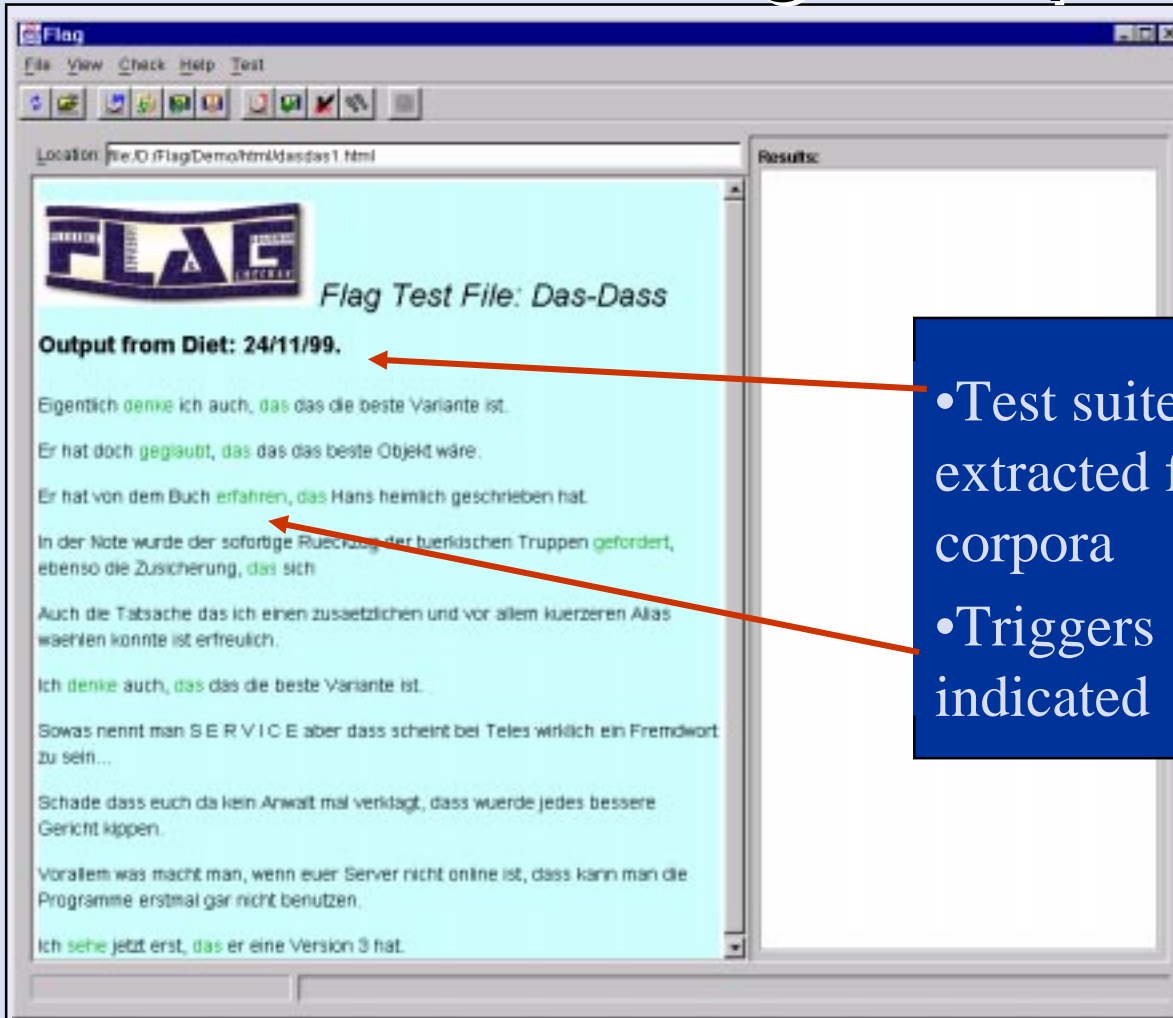
1. Erblasser [1] - 576-30/97-48841
2. Erblasser - ACH 1994-186638
3. Erblasser - ACH 1997-309021
4. Erblasser - FIS 1979-189136
5. Erblasser in der ACH - ACH-2225-48841

Document | Next | HitList | Query | Options | Feedback

German	Term	Erblasser [1]
English	Term	testator [1], deceased [2], legislator [3], the deceased [4]
Danish	Term	arvsløder
Spanish	Term	testador, causante
French	Term	testateur [1], de cujus [2], testateur [3]
	Phraseology	la succession comporte la possibilité pour l'époux d'exercer les droits de de cujus
Italian	Term	testatore, de cuius testis
Deutsch	Term	erbläßer [1], de cujus [2], de cujus [3], de cuius [4]
	Phraseology	plutöverrädding snarar de uppläsa vörk arvläsa gäst arslar om privata testamen. Zin- ning de de cujus by testament eras arvläsa
Portuguese	Term	"de cujus", testador
Finnish	Term	testamentiläijä
Swedish	Term	arvsläsa, testator



Development of error checking components I



- Test suites extracted from corpora
- Triggers indicated



Development of error checking components II

Debug Window

Meines Wissens nach wurde in den Verträgen lediglich die Entschädigungsfrage offengehalten .

Rules Full trace

Description *mWn* from 0 to 3

TRIGGER
50; Body: {@meines##meines, @wissens##wissens, @nach##nach}
POS_EV
80; Body: {\$nach, @prep##prep}
POS_EV
80; Body: {\$nach, @vfin##vfin}
NEG_EV
80; Body: {\$nach, @dat_obj##dat_obj}
NEG_EV
60; Body: {\$nach, -{@prep}|{@vfin}}, []* , @dat_obj##another_dat_obj
NEG_EV
40; Body: {@vfin##sep_verb, [-{@vfin}]* , \$meines}; Constraints:
[word({\$nach, \$sep_verb})]

Feature structures Configuration

- Sentence 6
 - 0: "Meines"
 - 1: "Wissens"
 - 2: "nach"
 - 3: "wurde"
 - MORPH
 - _top
 - LEMMA "werden"
 - READING
 - _top
 - CHUNK "1/1/W/_"
 - POS "VAFIN"
 - TOK "wurde"
 - 4: "in"
 - 5: "den"
 - 6: "Verträgen"
 - 7: "lediglich"
 - 8: "die"
 - 9: "Entschädigungsfrage"
 - 10: "offengehalten"
 - 11: "."

Environment for inspecting results