

Chunking—Practical Exercise

Chunking for NER

- Chunking, as we saw at the beginning, means finding parts of text
- This task is often called Named Entity Recognition (NER), in the context of finding person and organization names
- The same principle can be applied to any task that involves finding where things are located in text
 - For example, finding the noun phrases
 - Can you think of any others?

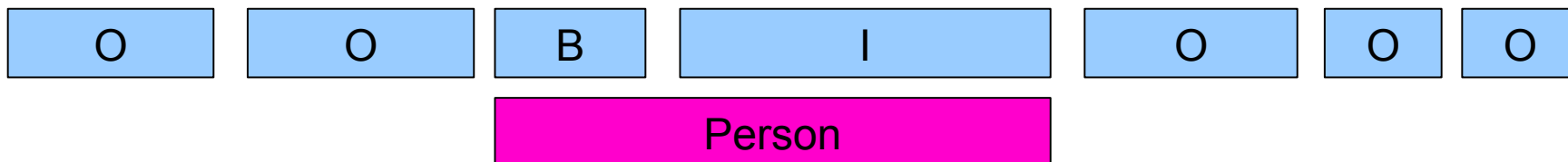
California Governor Arnold Schwarzenegger proposes deep cuts.

Person

Chunking for NER

- It's implemented as a twist on classification (everything is classification under the hood!)
- We achieve this in the Learning Framework by identifying which tokens are the beginning of a mention, which are the insides and which are the outsides ("BIO")
 - There are other schemes; the old Batch Learning PR used BE (beginnings and ends)
- You don't need to worry about the Bs, Is and Os; the Learning Framework will take care of all that for you! You just need a corpus annotated with entities

California Governor Arnold Schwarzenegger proposes deep cuts.



Chunking—Practical Exercise

- Materials for this exercise are in the folder called “chunking-hands-on”
- You might want to start by closing any applications and corpora from the previous exercise, so we have a fresh start

Finding UMLS Mentions using Chunking Training and Application PRs

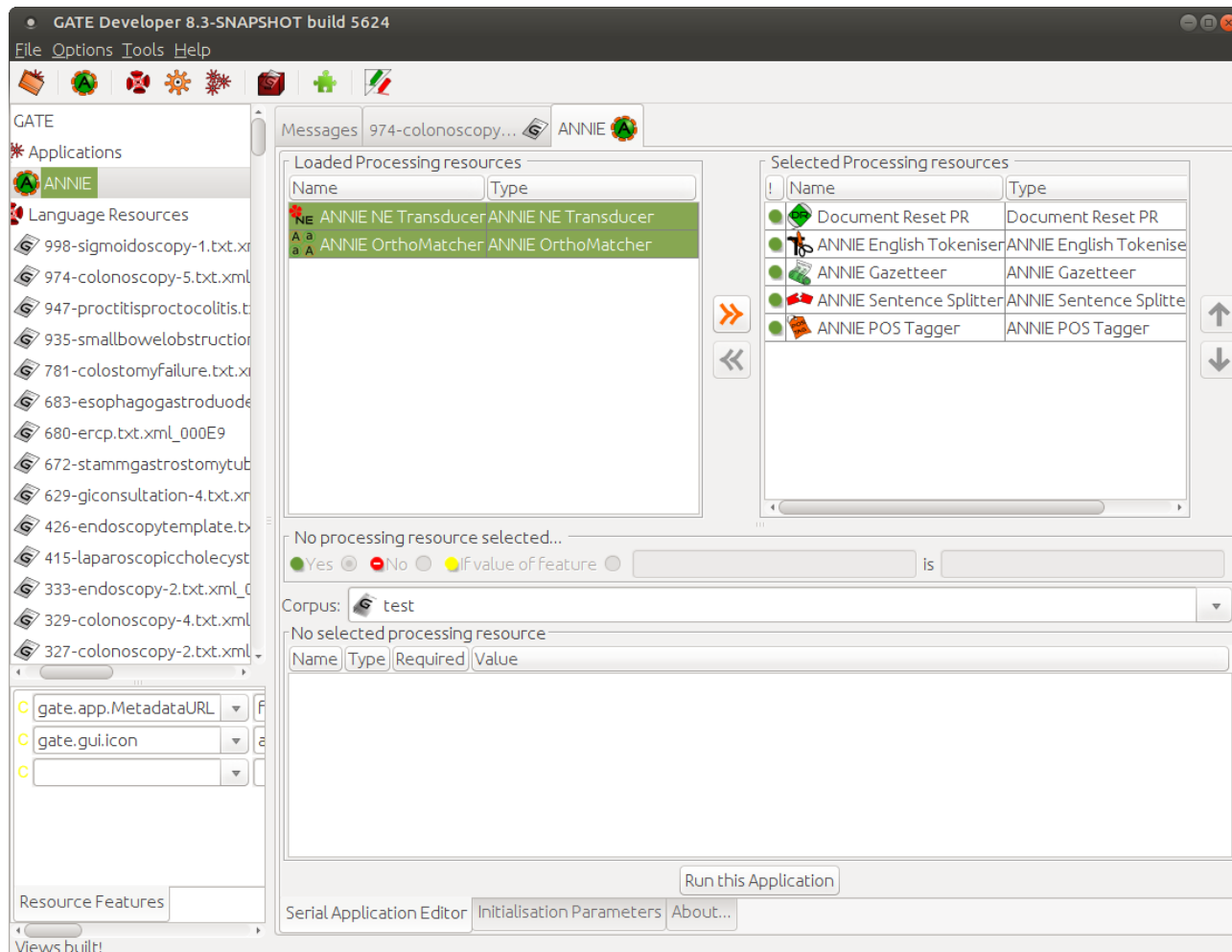
Load the corpus

- Create corpora for training and testing, with sensible names
- Populate them from the training and testing corpora you have in your chunking hands on materials
- Open a document and examine its annotations

Examining the corpus

- The corpus contains an annotation set called “Key”, which has been manually prepared
- Within this annotation set are annotations of types “Disease”, “Drug” and “Anatomy”

Creating the application



- As previously, if we run ANNIE on the corpus, we have more annotations to work with
- So start by loading ANNIE as the basis for your application
- Again, we don't need the NE transducer or orthomatcher

NER GATE application



GATE Developer 8.2-SNAPSHOT build 5490

File Options Tools Help

GATE

- Applications
 - ANNIE
- Language Resources
- Processing Resources
 - Annotation Set Transfer
 - LF_ApplyChunking 00031
 - LF_TrainChunking 00030
 - ANNIE OrthoMatcher
 - ANNIE NE Transducer
 - ANNIE POS Tagger
 - ANNIE Sentence Splitter
 - ANNIE Gazetteer
 - ANNIE English Tokeniser
 - Document Reset PR
- Datastores

Messages ANNIE

Loaded Processing resources

Name	Type
ANNIE NE Transducer	ANNIE NE Transducer
ANNIE OrthoMatcher	ANNIE OrthoMatcher
LF_ApplyChunking 00031	LF_ApplyChunking

Selected Processing resources

Name	Type
Document Reset PR	Document Reset PR
ANNIE English Tokeniser	ANNIE English Tokeniser
ANNIE Gazetteer	ANNIE Gazetteer
ANNIE Sentence Splitter	ANNIE Sentence Splitter
ANNIE POS Tagger	ANNIE POS Tagger
Annotation Set Transfer 00036	Annotation Set Transfer
LF_TrainChunking 00030	LF_TrainChunking

Run "Annotation Set Transfer 00036"?

☒ Yes ☐ No ☐ If value of feature is

Corpus: <none>

Runtime Parameters for the "Annotation Set Transfer 00036" Annotation Set Transfer:

Name	Type	Required	Value
annotationTypes	ArrayList		[]
copyAnnotations	Boolean	✓	false
inputASName	String		
outputASName	String		
tagASName	String		Original markups
textTagName	String		

Run this Application

Serial Application Editor Initialisation Parameters About...

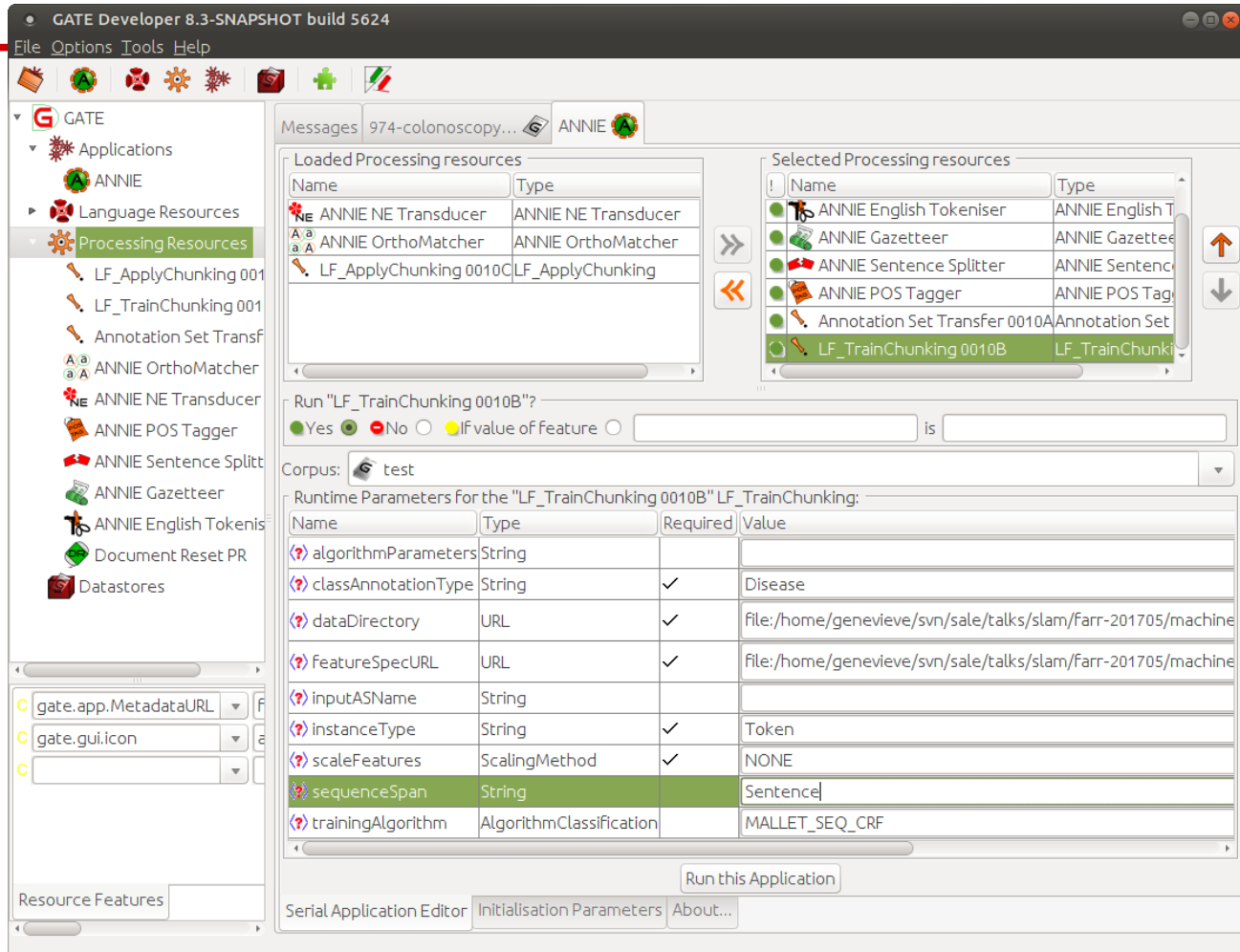
Annotation Set Transfer 00036 loaded in 0.001 seconds

- Again, we need an Annotation Set Transfer, so create and add one
- Then create both training and application chunking PRs
- Start by just adding the training one

Annotation Set Transfer

- We'll use the annotation set transfer to copy the Disease annotations up to the default annotation set, where we can learn them
- Go ahead and set up your AST now
- Be sure to copy them, not move them!

Chunking training parameters



The screenshot shows the GATE Developer interface with the 'LF_TrainChunking 0010B' resource selected. The 'Loaded Processing resources' list includes ANNE NE Transducer, ANNE OrthoMatcher, and LF_ApplyChunking. The 'Selected Processing resources' list includes ANNE English Tokeniser, ANNE Gazetteer, ANNE Sentence Splitter, ANNE POS Tagger, Annotation Set Transfer 0010A, and LF_TrainChunking 0010B.

Below the resource lists, the 'Run "LF_TrainChunking 0010B"?' dialog is shown with 'Yes' selected. The 'Corpus' is set to 'test'.

The 'Runtime Parameters for the "LF_TrainChunking 0010B" LF_TrainChunking:' table is displayed:

Name	Type	Required	Value
algorithmParameters	String		
classAnnotationType	String	✓	Disease
dataDirectory	URL	✓	file:/home/genevieve/svn/sale/talks/slam/farr-201705/machine
FeatureSpecURL	URL	✓	file:/home/genevieve/svn/sale/talks/slam/farr-201705/machine
inputASName	String		
instanceType	String	✓	Token
scaleFeatures	ScalingMethod	✓	NONE
sequenceSpan	String		Sentence
trainingAlgorithm	AlgorithmClassification		MALLET_SEQ_CRF

The 'Run this Application' button is visible at the bottom right of the configuration window.

- Let's look at the parameters for the training PR
- Instead of targetFeature, we have classAnnotationType

Chunking training parameters

- For classification, the class to learn is in a feature on the instance, is specified to the PR in the targetFeature parameter
- For chunking, the class to learn takes the form of an annotation type. In our case, our corpus is annotated with Person annotations that we are going to learn to locate
- This type to learn is indicated in the classAnnotationType parameter

Chunking training parameters

- Set the `classAnnotationType` now
- Set the `dataDirectory` to where you want to save your model, and set the `featureSpecURL` (there's a feature spec to get you started in the hands on materials)
- Set `instanceType`. What do you think it should be?

Sequence Spans

- sequenceSpan is only relevant when using sequence learners
- Sequence learners classify each instance in the span by making use of the others
- For example, a noun phrase might be more likely to follow a determiner than a preposition, or a disease name might be more likely to follow the words “diagnosed with”
- The Learning Framework offers the Conditional Random Fields sequence learner
- It might be good for finding diseases, so let's use it!
 - You don't have to use a sequence learner for chunking though
- What do you think would be a good sequence span?

Sequence Spans

- Sequence spans should be spans within which instance classes follow patterns
 - For example, grammatical rules apply to sequences of parts of speech
 - However, sentiment classifications of individual customer reviews don't form a meaningful sequence
- A sequence span shouldn't be longer than necessary
- Sentence would be a good span for our task
- Fortunately, ANNIE creates sentence annotations for us, so those are available to use
- **Set sequenceSpan to "Sentence"**

Feature Specification

```
<ML-CONFIG>
```

```
<ATTRIBUTE>  
<TYPE>Token</TYPE>  
<FEATURE>category</FEATURE>  
<DATATYPE>nominal</DATATYPE>  
</ATTRIBUTE>
```

```
<ATTRIBUTE>  
<TYPE>Token</TYPE>  
<FEATURE>kind</FEATURE>  
<DATATYPE>nominal</DATATYPE>  
</ATTRIBUTE>
```

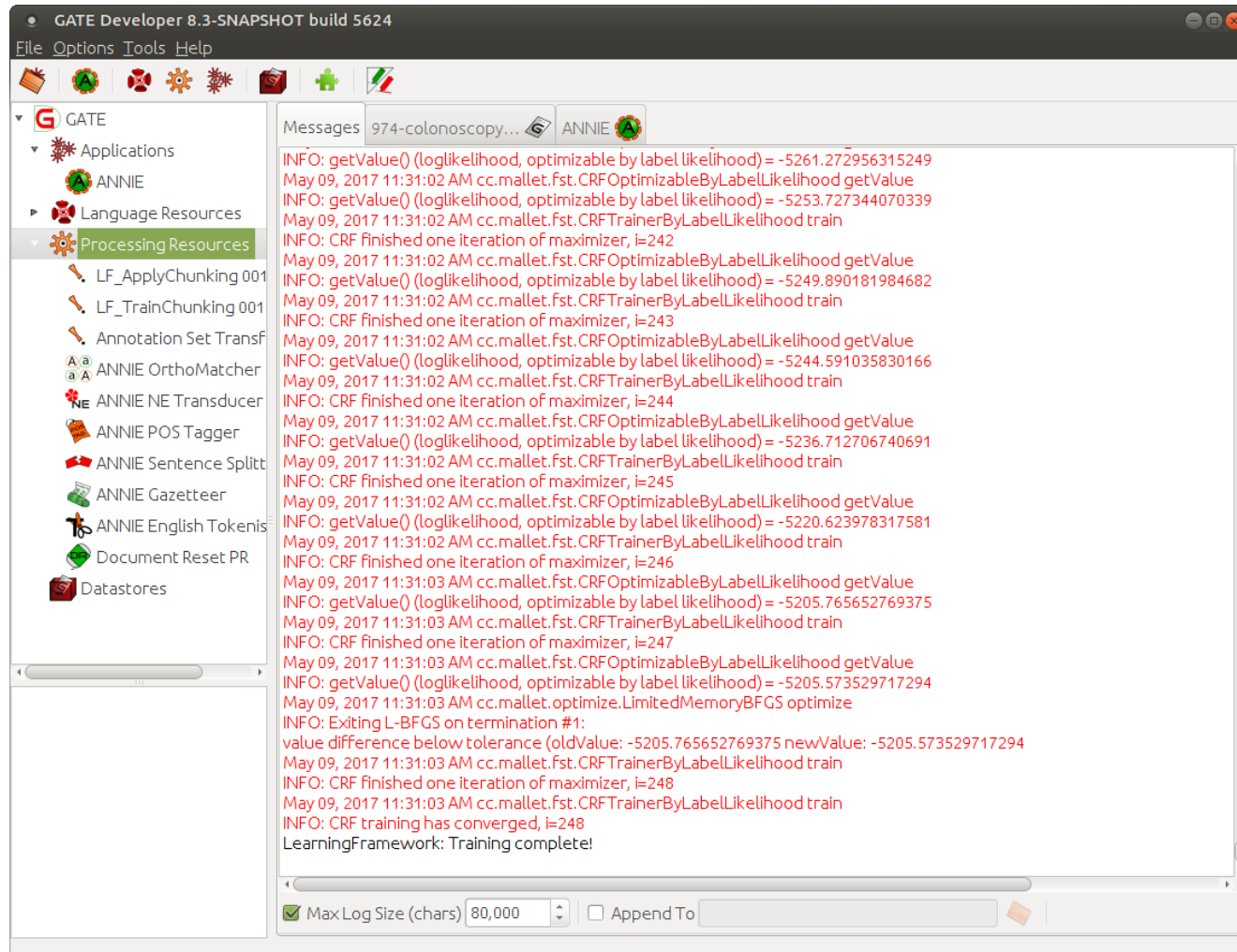
```
<ATTRIBUTE>  
<TYPE>Token</TYPE>  
<FEATURE>length</FEATURE>  
<DATATYPE>numeric</DATATYPE>  
</ATTRIBUTE>
```

```
<ATTRIBUTE>  
<TYPE>Token</TYPE>  
<FEATURE>orth</FEATURE>  
<DATATYPE>nominal</DATATYPE>  
</ATTRIBUTE>
```

```
<ATTRIBUTE>  
<TYPE>Token</TYPE>  
<FEATURE>string</FEATURE>  
<DATATYPE>nominal</DATATYPE>  
</ATTRIBUTE>
```

```
</ML-CONFIG>
```

- For this task, we are using attribute features
- These allow us to take features from the instance annotations or others that are co-located with them
- We specify type, feature and datatype
- Attribute features also can be taken from instances nearby
- That's a bit less useful with a sequence learner though—why?



- Make sure you have selected the training corpus
- Run the application!

Chunking application parameters

- Now move the training PR out of the application and add the application PR
- You can take the annotation set transfer out too
- The application PR doesn't have a targetFeature parameter like the classification application PR did
- You don't need to tell it what type to create because the model knows it from training!
- Set dataDirectory to the location where you told the training PR to put the model
- Set the sequence span

Applying

- Now run this on the test corpus

GATE Developer 8.3-SNAPSHOT build 5624

File Options Tools Help

Messages 974-colonoscopy... ANNIE

Loaded Processing resources

Name	Type
ANNIE NE Transducer	ANNIE NE Transducer
ANNIE OrthoMatcher	ANNIE OrthoMatcher
Annotation Set Transfer 0010A	Annotation Set Transfer
LF_TrainChunking 0010B	LF_TrainChunking

Selected Processing resources

Name	Type
Document Reset PR	Document Reset PR
ANNIE English Tokeniser	ANNIE English Tokeniser
ANNIE Gazetteer	ANNIE Gazetteer
ANNIE Sentence Splitter	ANNIE Sentence Splitter
ANNIE POS Tagger	ANNIE POS Tagger
LF_ApplyChunking 0010C	LF_ApplyChunking

Run "LF_ApplyChunking 0010C"?

☒ Yes ☐ No ☐ If value of feature is

Corpus: test

Runtime Parameters for the "LF_ApplyChunking 0010C" LF_ApplyChunking:

Name	Type	Required	Value
algorithmParameters	String		
confidenceThreshold	Double	✓	0.0
dataDirectory	URL	✓	file:/home/genevieve/svn/sale/talks/slam/Farr-201705/machine-learning/chun
inputASName	String		
instanceType	String	✓	Token
outputASName	String		LearningFramework
sequenceSpan	String		Sentence

Run this Application

Serial Application Editor Initialisation Parameters About...

Resource Features

gate.app.MetadataURL
gate.gui.icon

ANNIE run in 1.244 seconds

Chunking—Evaluation using Corpus QA

Chunking Evaluation

- For classification, each response is simply right or wrong
- For NER, there are more ways to be wrong
 - Fewer or more mentions than there really are, or you can overlap
- So we need different metrics

What are precision, recall and F1?

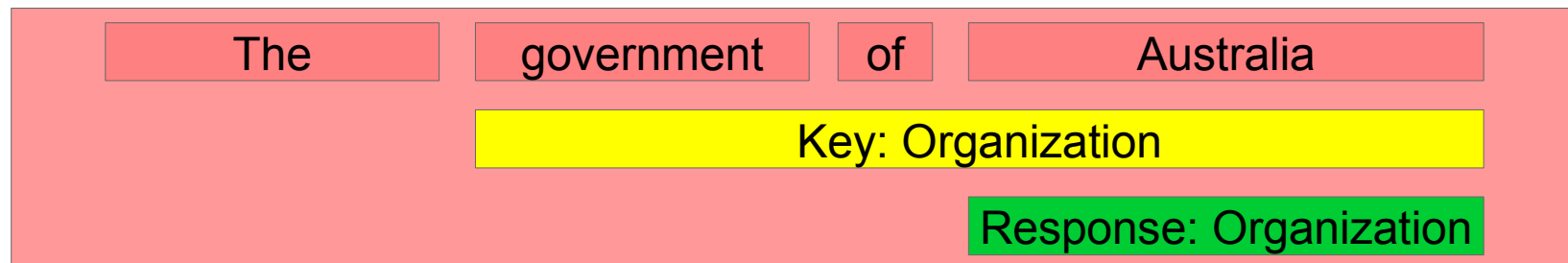
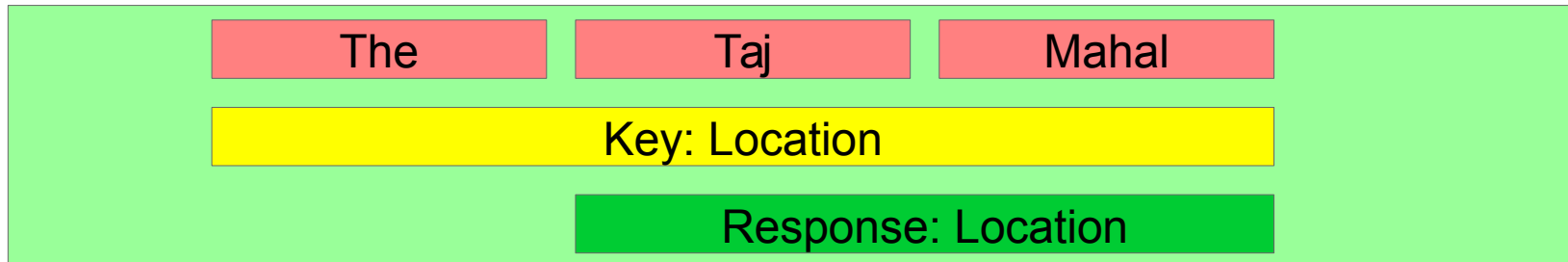
- Precision: what proportion of our automatic annotations were correct?
- Recall: what proportion of the correct annotations did our automatic tool create?
- $P = \text{correct} / (\text{correct} + \text{spurious}) = \text{tp} / (\text{tp} + \text{fp})$
- $R = \text{correct} / (\text{correct} + \text{missing}) = \text{tp} / (\text{tp} + \text{fn})$
- where tp = true positives, fp = false positives, fn = false negatives

What are precision, recall and F1?

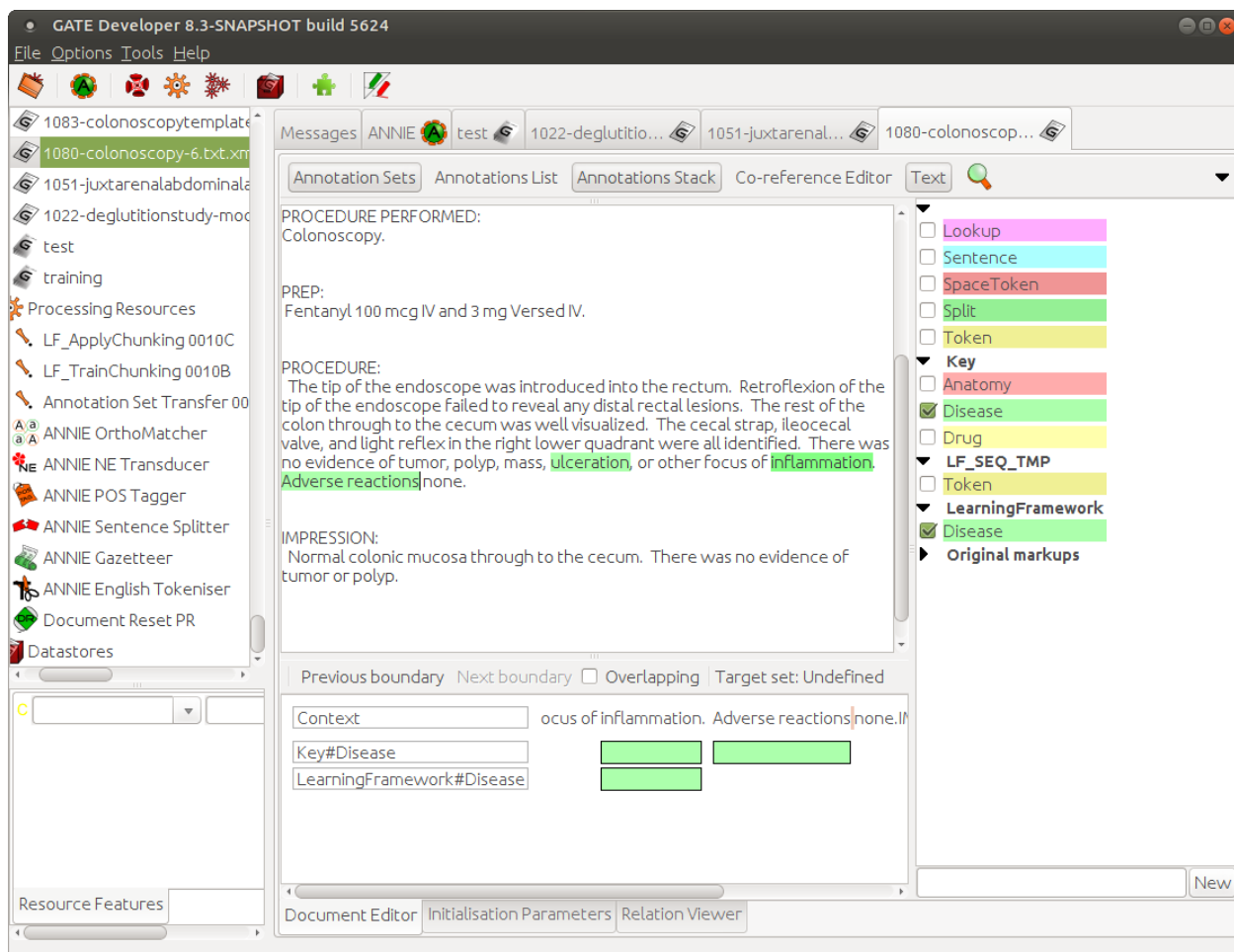
- F-score is an amalgam of the two measures
- $F_{\beta} = (1 + \beta^2)PR / (\beta^2 P + R)$
 - The equally balanced F1 ($\beta = 1$) is the most common F-measure
 - $F1 = 2PR / (P + R)$

Strict and Lenient

- “Strict” means we count an annotation as correct only if it has the same span as the gold standard annotation
- Lenient means we allow an annotation that overlaps to be correct, even if it isn't a perfect span match
- Which do you think is the right way to do it?



Examining the results of application



The screenshot displays the GATE Developer 8.3-SNAPSHOT build 5624 interface. The main window shows a document titled "1080-colonoscopy-6.txt.xml" with the following text:

PROCEDURE PERFORMED:
Colonoscopy.

PREP:
Fentanyl 100 mcg IV and 3 mg Versed IV.

PROCEDURE:
The tip of the endoscope was introduced into the rectum. Retroflexion of the tip of the endoscope failed to reveal any distal rectal lesions. The rest of the colon through to the cecum was well visualized. The cecal strap, ileocecal valve, and light reflex in the right lower quadrant were all identified. There was no evidence of tumor, polyp, mass, ulceration, or other focus of inflammation. Adverse reactions none.

IMPRESSION:
Normal colonic mucosa through to the cecum. There was no evidence of tumor or polyp.

The Annotations Stack on the right shows the following annotations:

- Lookup
- Sentence
- SpaceToken
- Split
- Token
- Key
- Anatomy
- Disease
- Drug
- LF_SEQ_TMP
- Token
- LearningFramework
- Disease
- Original markups

The Annotations Stack also shows the following annotations:

- Lookup
- Sentence
- SpaceToken
- Split
- Token
- Key
- Anatomy
- Disease
- Drug
- LF_SEQ_TMP
- Token
- LearningFramework
- Disease
- Original markups

The Annotations Stack also shows the following annotations:

- Lookup
- Sentence
- SpaceToken
- Split
- Token
- Key
- Anatomy
- Disease
- Drug
- LF_SEQ_TMP
- Token
- LearningFramework
- Disease
- Original markups

Examine a document from the test corpus

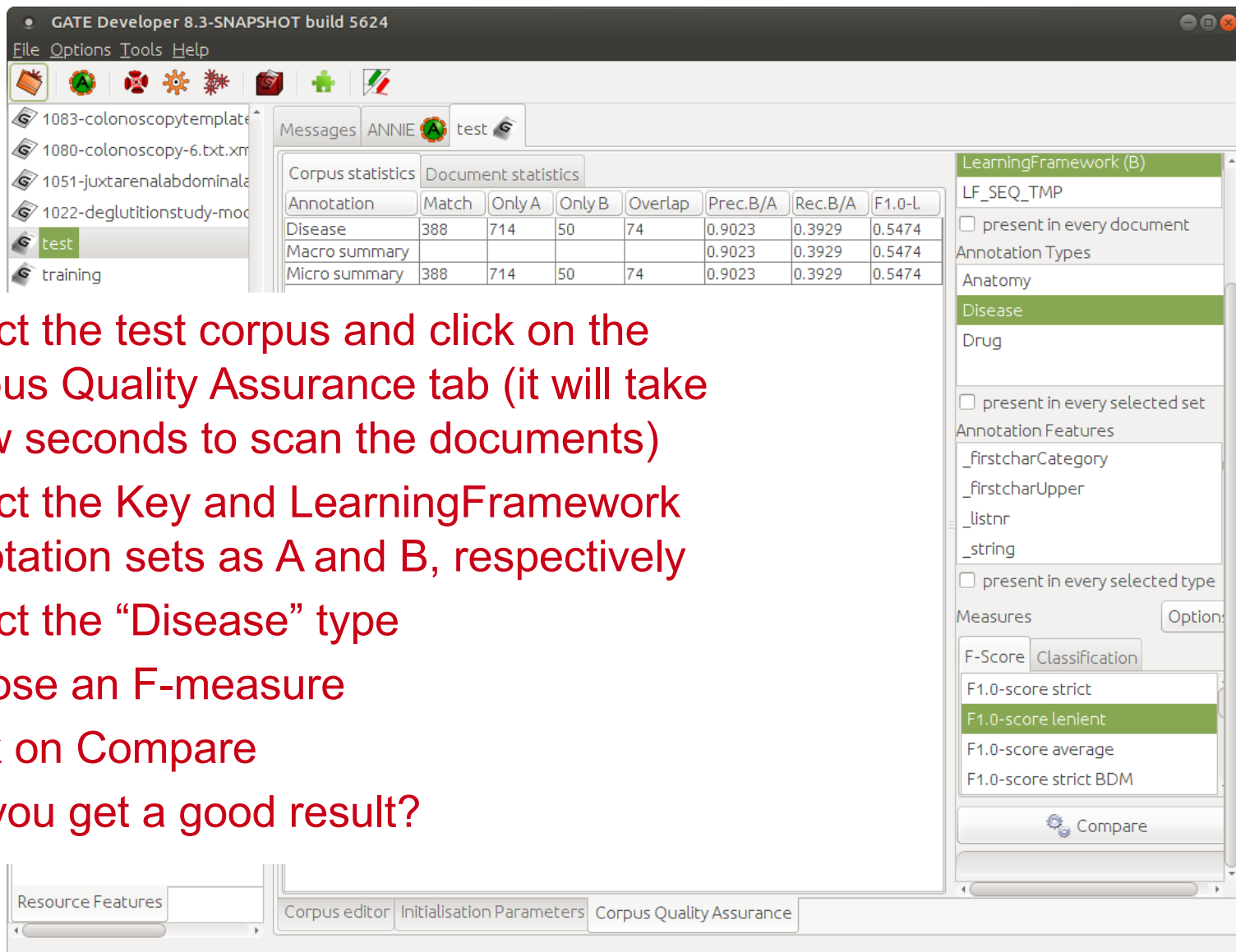
You should have a new “LearningFramework” AS with Person annotations

The original Person annotations (in the Key AS) are similar but not always identical!

The Annotations Stack is good for comparing them

How similar do they appear to be? Do you think you will get a good result?

Comparing the Sets with Corpus QA



GATE Developer 8.3-SNAPSHOT build 5624

File Options Tools Help

Messages ANNE test

Corpus statistics Document statistics

Annotation	Match	Only A	Only B	Overlap	Prec.B/A	Rec.B/A	F1.0-L
Disease	388	714	50	74	0.9023	0.3929	0.5474
Macro summary					0.9023	0.3929	0.5474
Micro summary	388	714	50	74	0.9023	0.3929	0.5474

LearningFramework (B)

LF_SEQ_TMP

☐ present in every document

Annotation Types

Anatomy

Disease

Drug

☐ present in every selected set

Annotation Features

_firstcharCategory

_firstcharUpper

_listnr

_string

☐ present in every selected type

Measures

F-Score Classification

F1.0-score strict

F1.0-score lenient

F1.0-score average

F1.0-score strict BDM

Compare

Resource Features

Corpus editor Initialisation Parameters Corpus Quality Assurance

- Select the test corpus and click on the Corpus Quality Assurance tab (it will take a few seconds to scan the documents)
- Select the Key and LearningFramework annotation sets as A and B, respectively
- Select the “Disease” type
- Choose an F-measure
- Click on Compare
- Did you get a good result?

Using Annotation Diff to examine performance

Annotation Difference

Key doc: 1305-esophagogastrodu... Key set: Key Type: Disease Weight: 1.0

Resp. doc: 1305-esophagogastrodu... Resp. set: LearningFram... Features: ☐ all ☐ some ☒ none

Compare

Start	End	Key	Features	=?	Start	End	Response	Features
207	221	abdominal.pain	{TUI=T184, Negation=...STY=Sign or Symptom}	=	207	221	abdominal.pain	{LF_confidence=0.843..
652	667	Crohn's-disease	{LABELVOCABS=CSP,MSH...Disease or Synd...	=	652	667	Crohn's-disease	{LF_confidence=0.613..
2992	3001	condition	{Experiencer=Patient...Disease or Syndrome}	=	2992	3001	condition	{LF_confidence=0.867..
3332	3336	PLAN	{Experiencer=Patient...Disease or Syndrome}	=	3332	3336	PLAN	{LF_confidence=0.902..
594	609	Crohn's-disease	{Negation=Affirmed, ...Disease or Syndrome}	=	594	609	Crohn's-disease	{LF_confidence=0.613..
753	766	COMPLICATIONS	{Negation=Affirmed, ...Pathologic Function}	=	753	766	COMPLICATIONS	{LF_confidence=0.697..
3313	3328	Crohn's-disease	{PREF=Crohn's, Negat...nticnetwork/id/T047}	=	3313	3328	Crohn's-disease	{LF_confidence=0.613..
4119	4128	stricture	{PREF=Stenoses, Nega...Pathologic Function}	~	4119	4138	stricture-formation	{LF_confidence=0.670..
3095	3110	aphthous-ulcers	{language=, Experien...Disease or Syndrome}	-?				
1451	1458	Crohn's	{PREF=Crohn's, Negat...Disease or Syndrome}	-?				
3228	3234	ulcers	{Negation=Affirmed, ...Disease or Syndrome}	-?				
4003	4013	strictures	{language=, Negation...Pathologic Function}	-?				
2638	2653	aphthous-ulcers	{Negation=Affirmed, ... Temporality=Recent}	-?				
2337	2352	multiple-ulcers	{LABELVOCABS=SNOMEDC...Pathologic Fu...	-?				
3590	3602	malnutrition	{inst=C0162429, Nega...Disease or Syndrome}	-?				
3158	3173	aphthous-ulcers	{language=, Negation...Disease or Syndrome}	-?				
4265	4272	Crohn's	{PREF=Crohn's, Negat...Disease or Syndrome}	-?				
3696	3704	syndrome	{TUI=T047, Negation=...Disease or Syndrome}	-?				
				?	400	416	gastrointestinal	{LF_confidence=0.521..
				?	253	280	recurrent-abscess-formation	{LF_confidence=0.572..

20 pairings have been found (0 annotations are hidden)

Correct: 7 Recall Precision F-measure

Partially correct: 1 Strict: 0.39 0.70 0.50

Missing: 10 Lenient: 0.44 0.80 0.57

False positives: 2 Average: 0.42 0.75 0.54

Statistics Adjudication

Switch to the
“Document
statistics” tab

Choose a
document

Click on the
Annotation Diff
icon



(The Annotation
Diff icon is at the
top right of the
Corpus QA
sidebar)

What kind of
mistakes did your
application make?

Using Annotation Diff...

- “Correct”: the response annotation has the right feature and span
- “Partially correct”: response has the right feature and overlapping but not exactly matched span; this counts as correct in the “lenient” scoring
- “Missing”: key annotation+feature is missing from the response (a.k.a. “false negative”)
- “False positive”: response annotation+feature shouldn't be there (a.k.a. “spurious”)

Classification Evaluation PR for Chunking?

- We didn't use a Learning Framework evaluation PR for this chunking task
- What do you think would happen if you used the Classification Evaluation PR to do a chunking problem?
- It would work! It would evaluate the accuracy of the system in correctly identifying beginnings, insides and outsides
- However, it wouldn't tell you much about how well you did finding named entities
 - There are so many outsides that you can get a high score just by saying everything is an outside!
- You could use it to tune parameters if you wanted, though

Exercise 1—Improving the result

- Again, see if you can improve your result
- Tips and ideas:
 - Use the classification evaluation PR to tune algorithm and parameters (this won't work for CRF)
 - Some algorithms can take a while to train*. Switch to the messages tab and run the application by right-clicking on it in the resources pane—that way you can see what is happening while you wait. Once training starts you can't interrupt!
 - Make notes of the results while you are tuning
 - For non-sequence algorithms you must remove the sequenceSpan parameter
 - Experiment with confidence thresholds?

*Fast algos: CRF, MaxEnt, Naive Bayes EM, Winnow, Balanced Winnow, Decision Tree
Slow algos (don't do it!!): C45
Quite slow: LibSVM

Exercise 2

-
- Try to learn different entity types