

Introduction:

Requirements for a Digital Ecosystem

Marco Büchler
Göttingen Centre for Digital Humanities

Lyon, France
3rd June 2014

ACID for the eHumanities

ACID for the eHumanities

Acceptance by scholars and computer scientists/developers for what we do?

Complexity of needs in the Humanities

Interoperability between data and tools

Diversity of natural language texts

Diversity by Re-use Types

Wisdom Quotation Wit Law Saw Verse Parole

Joke Quip Punch Line Platitude Proverb Rant

Slogan Palindrom Meme Mantra Maxim

Sententiae Motto Loanword Koan Legend

Fact Phraseme Idiom Epigram Definition Edition

Paroimia Gnome Bonmot Battle Cry Cliche

Simile Metaphor Ephithet Abstract Adage

Template Pangram Epitome Anagram Flowery Phrase

Triusm Parable Equation Aphorism Apophtegm

- **Stability (yellow):** syntactic vs. semantic

- **Purpose (green)**

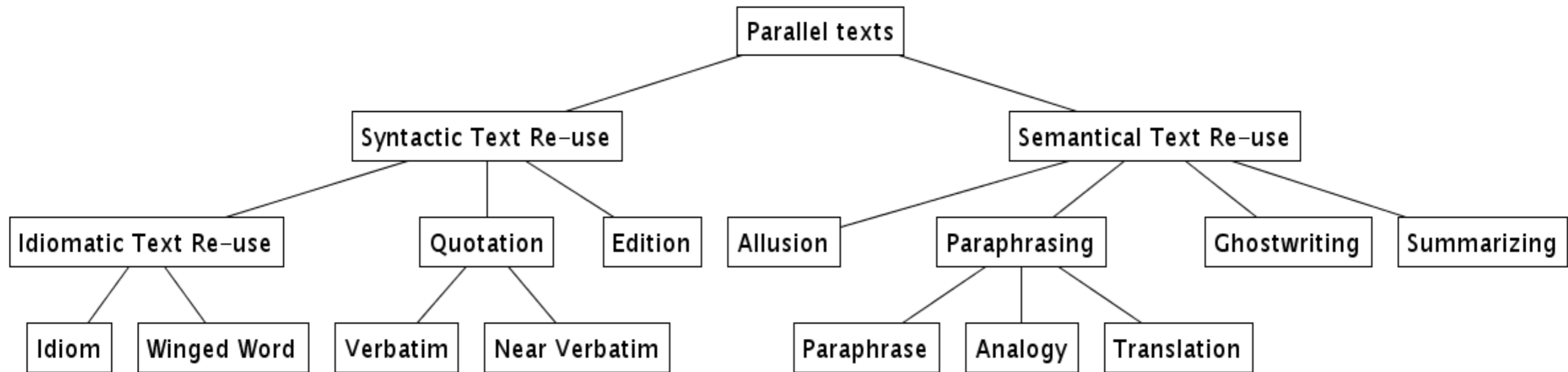
- **Size of Text Re-use (blue)**

- **Literary classification (light blue)**

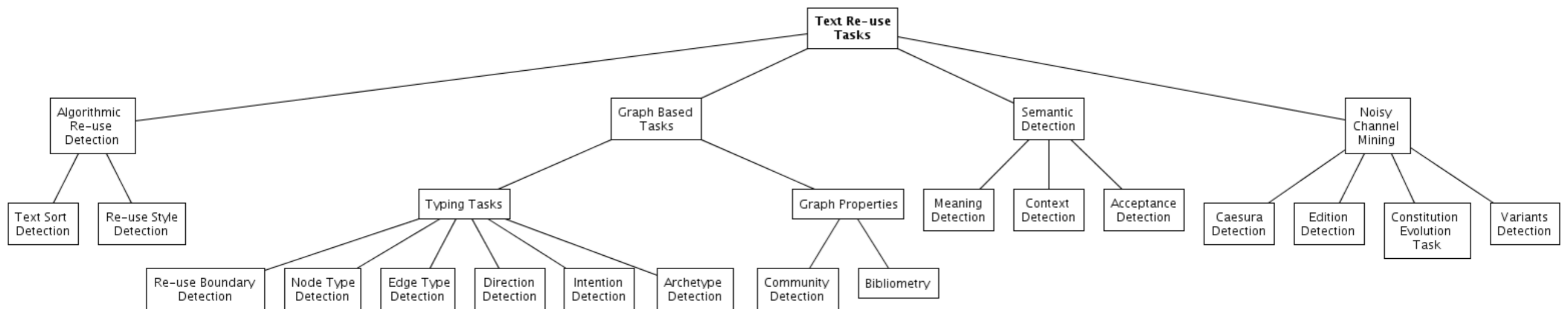
- **Degree of distribution (lila)**

- **Written and oral transmission**

Diversity by different *re-use* types

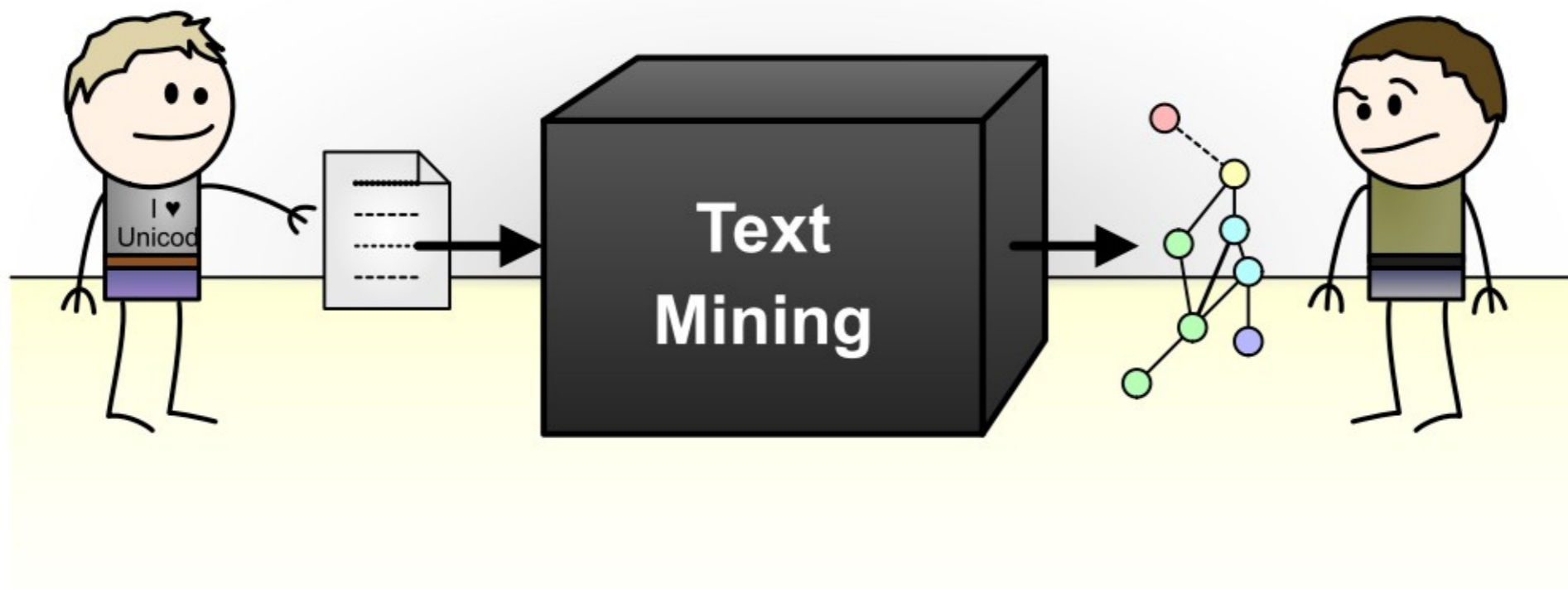


Complexity of needs in the Humanities



Acceptance of Language Models

Interdisciplinary collaborations: The problem!



Acceptance of Language Models

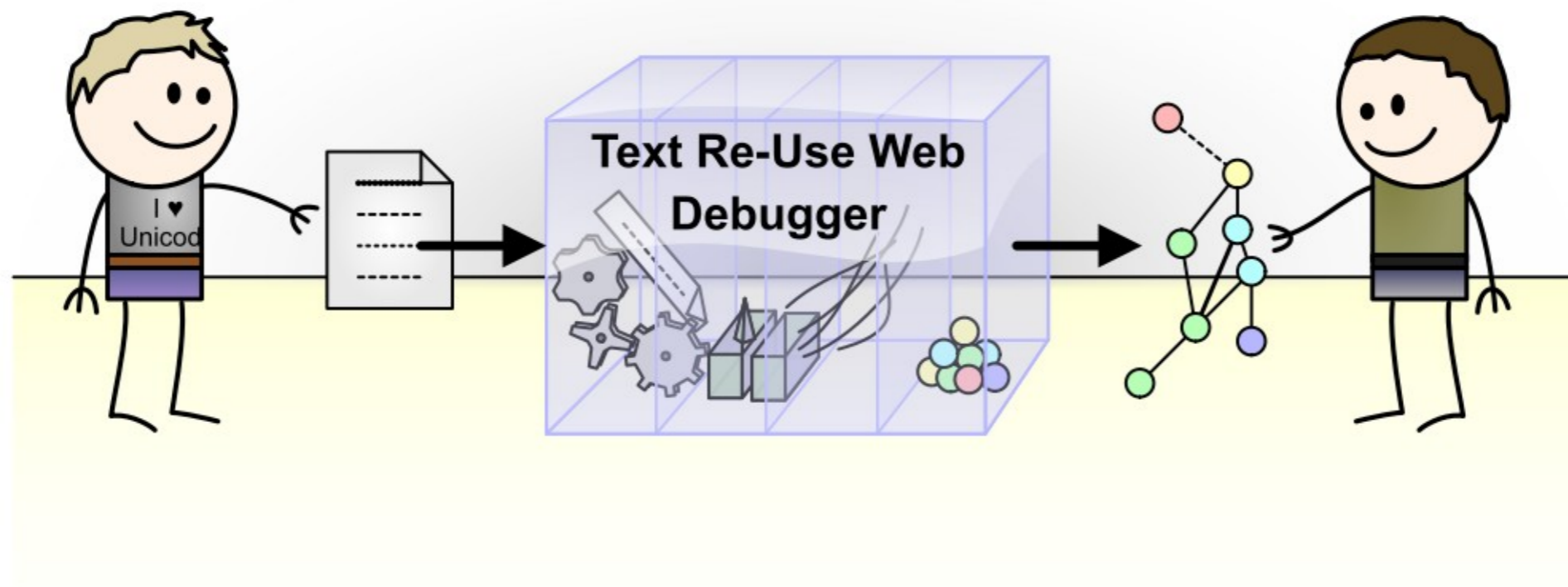
Change of view for understanding a Humanist's point



How to be accepted by humanists if text mining is a black box, that can't be looked in?

Acceptance of Language Models

Change of view for understanding a Humanist's point



Transparency: How to provide user-friendly insights into complex mining techniques and machine learning?

Acceptance of Language Models

Text Re-use Web Debugger: Starting point

Step 0: Searching

Please select a Corpus:*

bible

Please select the number of displayed sentences:

20

Input the Word you are searching for:*

God

Fields with * are necessary

Trace

In the beginning God created the heavens and the earth.

[Trace](#)

And the earth was waste and void; and darkness was upon the face of the deep: and the Spirit of God moved upon the face of the waters.

[Trace](#)

And God said, Let there be light: and there was light.

[Trace](#)

And God saw the light, that it was good: and God divided the light from the darkness.

[Trace](#)

And God called the light Day, and the darkness he called Night. And there was evening and there was morning, one day.

[Trace](#)

And God said, Let there be a firmament in the midst of the waters, and let it divide the waters from the waters.

[Trace](#)

And God made the firmament, and divided the waters which were under the firmament from the waters which were above the firmament: and it was so.

[Trace](#)

And God called the firmament Heaven. And there was evening and there was morning, a second day.

[Trace](#)

And God said, Let the waters under the heavens be gathered together unto one place, and let the dry land appear: and it was so.

[Trace](#)

And God called the dry land Earth; and the gathering together of the waters called he Seas: and God saw that it was good.

[Trace](#)

And God said, Let the earth put forth grass, herbs yielding seed, and fruit-trees bearing fruit after their kind, wherein is the seed thereof, upon the earth: and it was so.

[Trace](#)

And the earth brought forth grass, herbs yielding seed after their kind, and trees bearing fruit, wherein is the seed thereof, after their kind: and God saw that it was good.

[Trace](#)

And God said, Let there be lights in the firmament of heaven to divide the day from the night; and let them be for signs, and for seasons, and for days and years:

[Trace](#)

And God made the two great lights; the greater light to rule the day, and the lesser light to rule the night: he made the stars also.

[Trace](#)

And God set them in the firmament of heaven to give light upon the earth,

[Trace](#)

and to rule over the day and over the night, and to divide the light from the darkness: and God saw that it was good.

[Trace](#)

And God said, Let the waters swarm with swarms of living creatures, and let birds fly above the earth in the open firmament of heaven.

[Trace](#)

And God created the great sea-monsters, and every living creature that moveth, wherewith the waters swarmed, after their kind, and every winged bird after its kind: and God saw that it was good.

[Trace](#)

And God blessed them, saying, Be fruitful, and multiply, and fill the waters in the seas, and let birds multiply on the earth.

[Trace](#)

And God said, Let the earth bring forth living creatures after their kind, cattle, and creeping things, and beasts of the earth after their kind: and it was so.

[Trace](#)

prev 0 1 2 3 4 5 6 ... 1146 next

Acceptance of Language Models

Text Re-use Web Debugger: Preprocessing

▣ Step 0: Searching

▣ Step 1: Preprocessing

Please select a preprocessing strategy:

01:02-WLP:lem=true_syn=false_ssim=false_redwo=false:ngam=5:LLR=true_toLC=true_rDia=false_w2wl=false:wlt=5

change

Unprocessed Sentence:

In the beginning God created the heavens and the earth.

Preprocessed Sentence:

in the begin god create the heaven and the earth .

correct

Your correction for the processed sentence:

in the begin god create the heaven and the earth .

Your comment:

submit changes

Other users preference

No users have suggested a change in the preprocessing level

next Level

Acceptance of Language Models

Text Re-use Web Debugger: Featuring

▣ Step 0: Searching

▣ Step 1: Preprocessing

▣ Step 2: Featuring

Please select a training strategy:

Preprocessed sentence: in the begin god create the heaven and the earth .

Position	Feature
0	in the
1	the begin

Position	Feature
2	begin god
3	god create

Position	Feature
4	create the
5	the heaven

Position	Feature
6	heaven and
7	and the

Position	Feature
8	the earth
9	earth .

Acceptance of Language Models

Text Re-use Web Debugger: Selection

Step 3: Selecting

Please select a selecting strategy:

Agenda

word = This word belongs to the fingerprint

word = This word originally doesn't belong to the fingerprint but was selected by the user to belong to the fingerprint

word = This word doesn't belong to the fingerprint

word = This word originally belonged to the fingerprint but was selected by the user to not belong to the fingerprint

initial configuration: in the the begin begin god god create create the the heaven heaven and and the the earth earth .

current configuration: in the the begin begin god god create create the the heaven heaven and and the the earth earth .

selected features

<-->

not selected features

in the
the begin
god create
the heaven
heaven and
and the
the earth
earth .

begin god
create the

Other users preference

Feature	users selected	users not selected
in the	0	1
the begin	1	0
begin god	1	0
god create	1	0
create the	0	1
the heaven	1	0
heaven and	1	0
and the	0	1
the earth	1	0
earth .	0	1

Statistics

Feature	Selected Features	Total number of features
in the	27114	32227
the begin	470	480
begin god	0	5
god create	27	45
create the	17	38
the heaven	1624	1695
heaven and	389	398
and the	31808	40850
the earth	4776	5222
earth .	1030	1040

Acceptance for the Infrastructure

- **Local repository vs. Distributed system?**

- Costs for integration / costs for updating of running projects

- **Access to text?**

- Web page / download / CTS server

- **Interoperability** between data and tools

- **Why do you see some projects as „fails“?**

- Missing requirements: formal reqs, missing „complexity & diversity“?

Requirements for a
Digital Ecosystem
are about social aspects
FIRST!!!