CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt **Date of investigation**: July 30, 2008

Book title: The Confessions of St. Augustine
Author: Charles Eliot is the editor (this is a translated version)
Year of Publication: 1909 this version
Place of Publication: P. F. Collier & son
ISBN number?: NA
Content/Disciplinary area: theory, history; this is the first document we have in the filed of linguistics where someone (St. Augustine) attempts to describe the language learning process – often quoted.

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	This book provided to Google by Stanford
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	This book has wonderful metadata – several of the subtitles and sections are hyperlinked in the Table of Contents (see Figure 1). This makes jumping to specific sections very easy for the researcher.
MARC record available?	NA
Can the book be searched through an OPAC? If so, which one?	No
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	NA

SCAN SEARCH

search paradigm settings (example: every 10 double-pages) scrolled through entire book.

Page(s):	readability/ quality of scan	anomalies?	other?	other?
2	sometimes, the opposite page's text shows up on scanned page. Here is an example – this text was even included in the hyperlink.	(see figure 2)		
11	scan is completely illegible	(see figure 3)		
81	the scan of page 81 is massively distorted			
209, 211	These page has great hyperlinked sections again			
277	the text at the bottom of this page is slightly distorted			
368		scanner's fingers visible on this page		

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	Comment (Is OCR search permitted): Y / N
language	successful recognition
express	successful recognition (yellow highlighting for all results!)
natural language	successful recognition
gestures	successful recognition
memory	successful recognition
utter	successful recognition
	Etc- all recognition perfectly successful

IMAGE ASSESSMENT

If the original text includes graphics/illustrations, are they included in the electronic version?	No images in this text
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	
Is the scan on a gray-scale or color scale?	
Quality – does it look like the real thing?	
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	
Other notes	

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

GENERAL OBSERVATIONS

Comment on ease of discovery in integrated interface (OPAC or better)	Very easy in Google Books
Overall quality of scan of this book	Excellent – this book also has wonderful hyperlinks
Ease of access to full text	Very
Ability to copy, quote, cite?	Yes
Ability to search?	Yes
Ability to download?	Yes
Ability to print?	Yes
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	Very
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Yes
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes
Issues with this book that would affect usefulness for scholars	None
Can this book be shared with other scholars/students?	Yes, via emailing of the URL
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	No missing items
Other notes	Pagination, overall scan quality of this book is very good. Page 11 is only problematic page.

Book URL:
http://books.google.com/books?id=zDgLAAAAIAAJ&printsec=frontcover&dq=confessions+of+st+augustine#PPA386,M1
Date accessed: Wednesday, July 30, 2008

Figure 1 Metadata: hyperlinked sections in this book

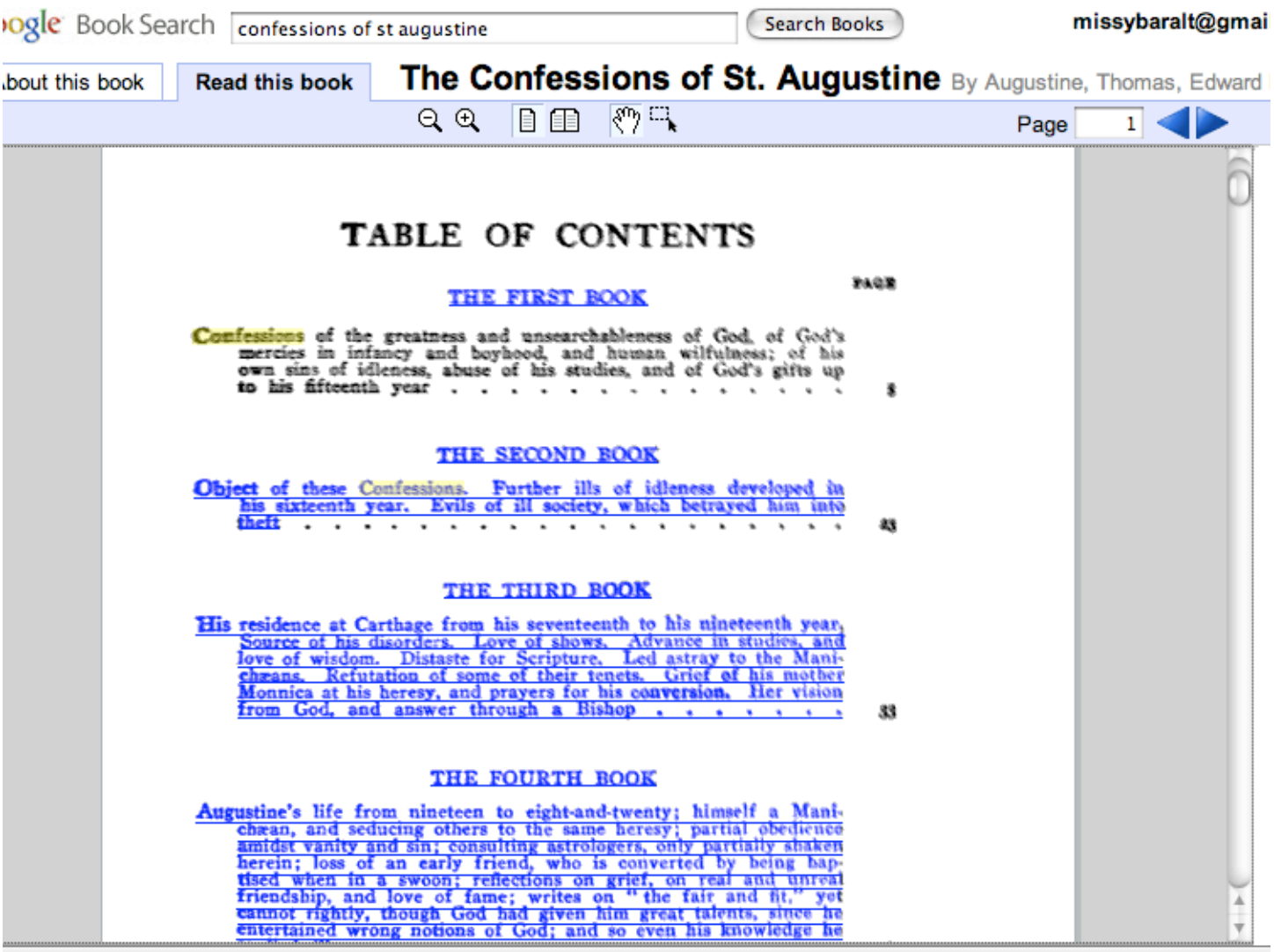
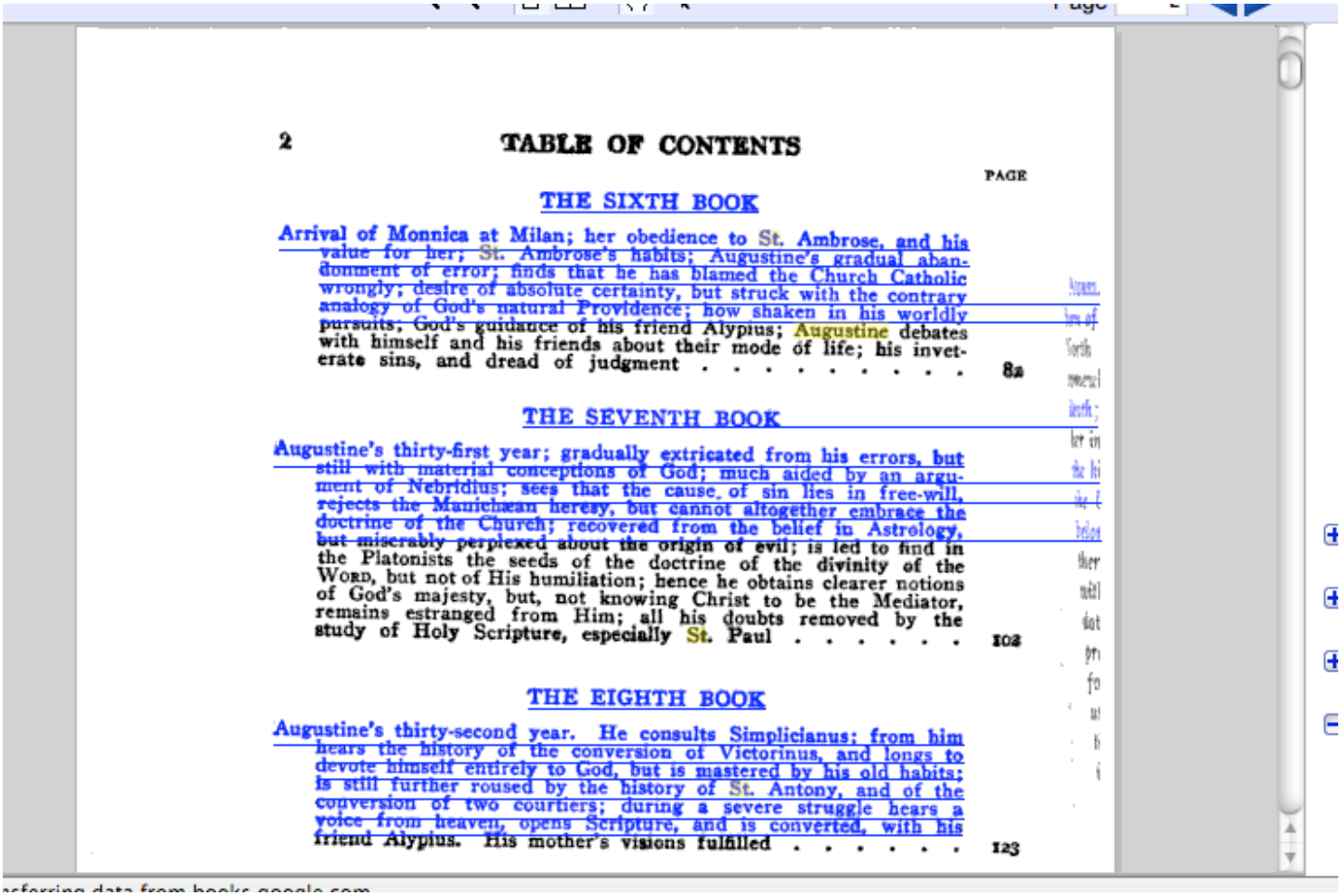


Figure 2 text from attached page



2 TABLE OF CONTENTS

PAGE

THE SIXTH BOOK

Arrival of Monnica at Milan; her obedience to St. Ambrose, and his value for her; St. Ambrose's habits; Augustine's gradual abandonment of error; finds that he has blamed the Church Catholic wrongly; desire of absolute certainty, but struck with the contrary analogy of God's natural Providence; how shaken in his worldly pursuits; God's guidance of his friend Alypius; Augustine debates with himself and his friends about their mode of life; his inveterate sins, and dread of judgment 82

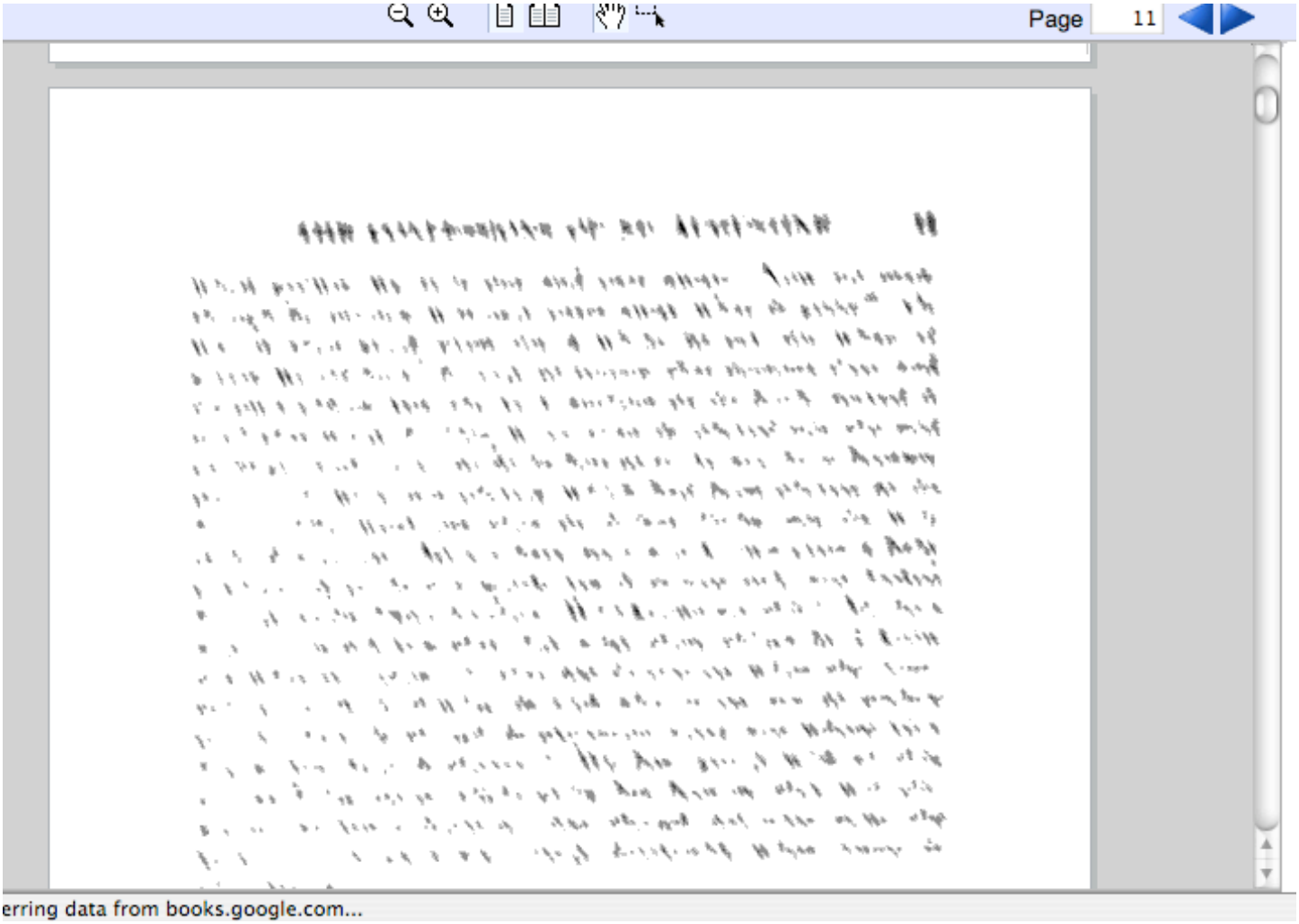
THE SEVENTH BOOK

Augustine's thirty-first year; gradually extricated from his errors, but still with material conceptions of God; much aided by an argument of Nebridius; sees that the cause of sin lies in free-will, rejects the Manichæan heresy, but cannot altogether embrace the doctrine of the Church; recovered from the belief in Astrology, but miserably perplexed about the origin of evil; is led to find in the Platonists the seeds of the doctrine of the divinity of the WORD, but not of His humiliation; hence he obtains clearer notions of God's majesty, but, not knowing Christ to be the Mediator, remains estranged from Him; all his doubts removed by the study of Holy Scripture, especially St. Paul 102

THE EIGHTH BOOK

Augustine's thirty-second year. He consults Simplicianus; from him hears the history of the conversion of Victorinus, and longs to devote himself entirely to God, but is mastered by his old habits; is still further roused by the history of St. Antony, and of the conversion of two courtiers; during a severe struggle hears a voice from heaven, opens Scripture, and is converted, with his friend Alypius. His mother's visions fulfilled 123

Figure 3 – page 11 (was obviously moved during scan, completely illegible)



CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt **Date of investigation**: August 1, 2008

I originally wanted to explore Bloomfield's book titled simply "Language" – that book provides a detailed account of the behaviorist perspective on language learning (published in 1933). It is only available for snippet view on Google Books, however. I decided to do
Book title: An Introduction to the Study of Language, available for full view.
Author: Leonard Bloomfield
Year of Publication: 1914
Place of Publication: H. Holt and Company
ISBN number?: NA
Content/Disciplinary area: language, linguistics

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	Original was provided by Harvard University
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Very easy in Google Books
MARC record available?	NA
Can the book be searched through an OPAC? If so, which one?	No
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	Google's 'about this book' link provides all of the necessary information – it is all accurate for this book. This book also has wonderful hyperlinks in the Table of Contents and in the subject index in the back, making it easy for the reader/researcher to jump to those sections of the book. (see Figure 1).

SCAN SEARCH

search paradigm settings (example: every 10 double-pages) Scanned through entire book

Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
173		a patron's personal checkout stickie is on this page and was included in the scan, preventing reading of the text behind the sticker on this page.	See Figure 2	
220		black circular figure covers some of text on bottom half of page		

No other anomalies

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	Comment (Is OCR search permitted): Y / N
pitch	successful recognition
□	not successful – this letter is used often in the phonetic description section
è, [è]	not successful – this letter is used often in the phonetic description section
ö	not successful – this letter is used often in the phonetic description section
∫	not successful – this letter is used often in the phonetic description section
morphology	successful recognition
syntax	successful recognition
linguistics	successful recognition
teaching	successful recognition
sanskrit	successful recognition

OCR recognition is successful except for any IPA [International Phonetic Association] symbols

IMAGE ASSESSMENT

If the original text includes graphics/illustrations, are they included in the electronic version?	No images in this text
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	
Is the scan on a gray-scale or color scale?	
Quality – does it look like the real thing?	
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming	

capabilities, filter options)

Other notes

GENERAL OBSERVATIONS

Comment on ease of discovery in integrated interface (OPAC or better)	Easily accessible in Google books
Overall quality of scan of this book	Excellent- perfect pagination, hyperlinks in table of contexts and subject sections
Ease of access to full text	Yes
Ability to copy, quote, cite?	Yes
Ability to search?	Yes
Ability to download?	Yes
Ability to print?	Yes
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	Great except for any IPA symbols –a problem for linguistics
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Yes
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes
Issues with this book that would affect usefulness for scholars	No searchability of IPA phonetic symbols
Can this book be shared with other scholars/students?	Yes, by emailing the URL
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	NA
Other notes	

Book URL:
http://books.google.com/books?id=YZkOAAAAYAAJ&printsec=titlepage&dq=intitle:Language+inauthor:Bloomfield&lr=&as_brr=0&source=gbs_summary_r&cad=0#PPR7,M1

Date accessed: August 1, 2008

Figure 1

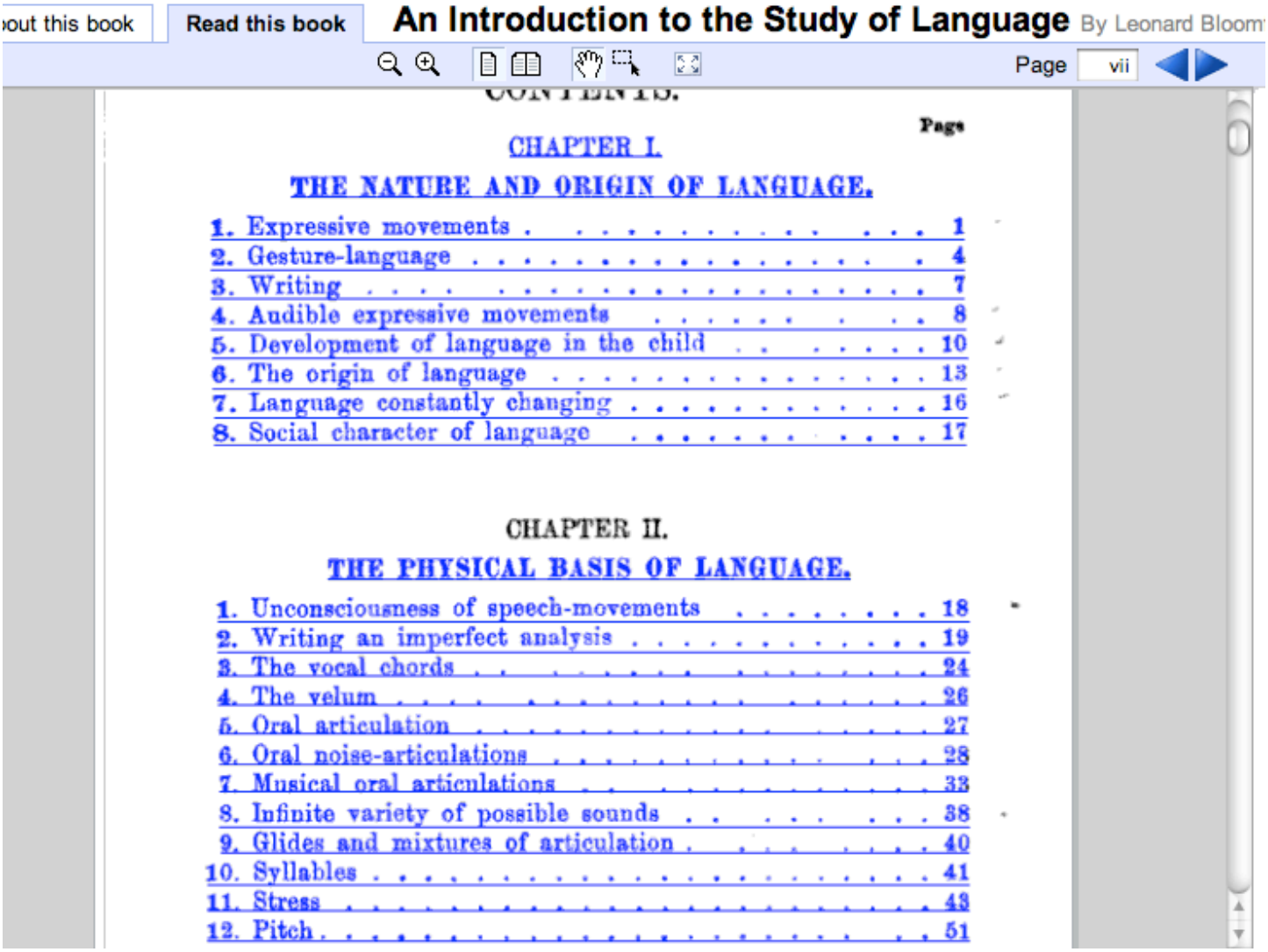
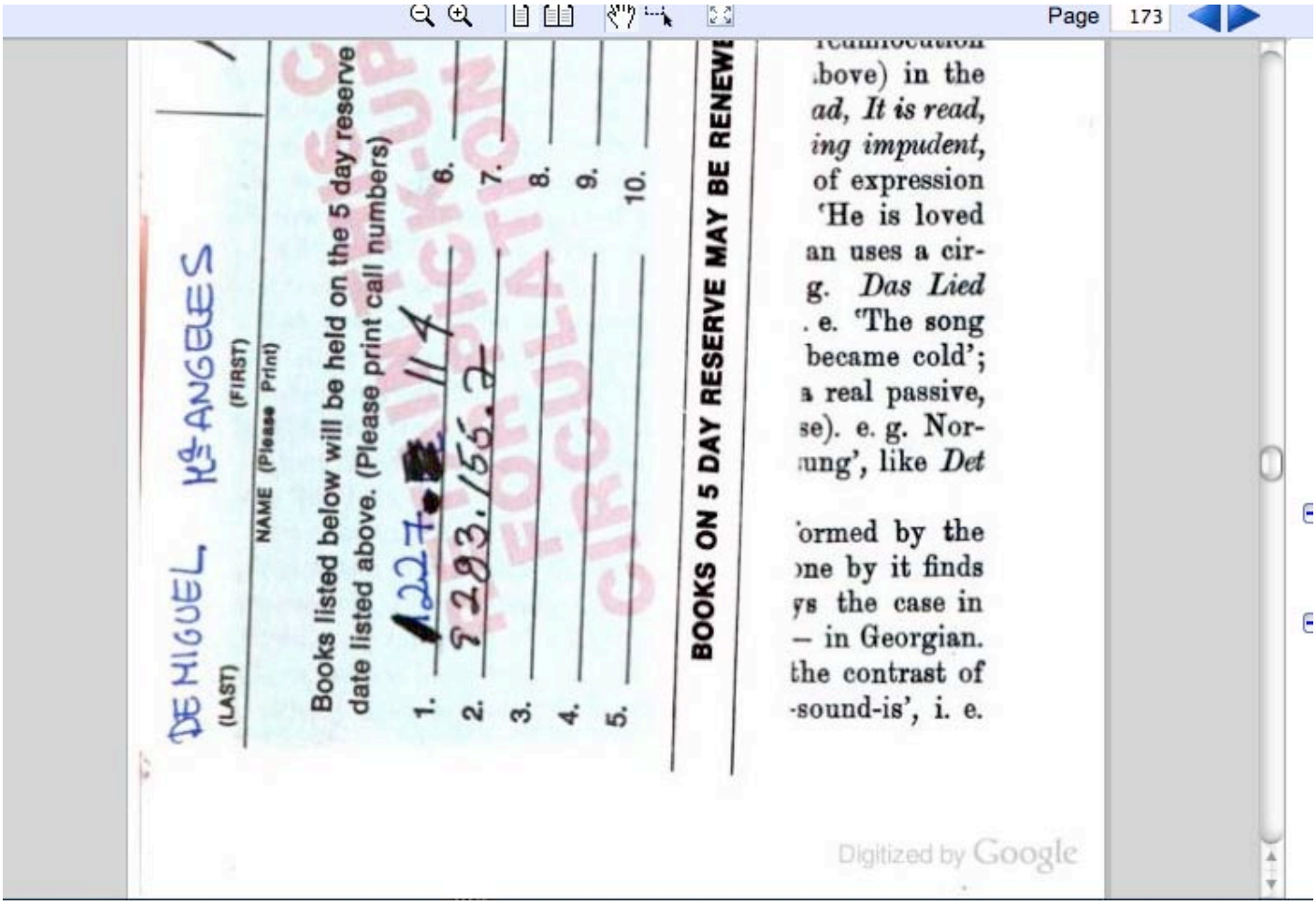

out this book
Read this book
An Introduction to the Study of Language
By Leonard Bloom
Page
vii
CONTENTS.
Page
CHAPTER I.
THE NATURE AND ORIGIN OF LANGUAGE.
1. Expressive movements 1
2. Gesture-language 4
3. Writing 7
4. Audible expressive movements 8
5. Development of language in the child 10
6. The origin of language 13
7. Language constantly changing 16
8. Social character of language 17
CHAPTER II.
THE PHYSICAL BASIS OF LANGUAGE.
1. Unconsciousness of speech-movements 18
2. Writing an imperfect analysis 19
3. The vocal chords 24
4. The velum 26
5. Oral articulation 27
6. Oral noise-articulations 28
7. Musical oral articulations 33
8. Infinite variety of possible sounds 38
9. Glides and mixtures of articulation 40
10. Syllables 41
11. Stress 43
12. Pitch 51

Figure 2
patron sticker on page 173 that was included in the scan – this anomaly prevents reading of the text on this page



CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt **Date of investigation**: July 27, 2008

Book title: The Childes Project: Tools for Analyzing Talk
Author: Brian MacWhinney
Year of Publication: 2000
Place of Publication: Lawrence Erlbaum Associates
ISBN number?: 0805829954
Content/Disciplinary area: linguistics, child language, language data analysis

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	This is a limited preview edition. However, even though it is limited view, I was very surprised to see the amount of detail that IS available. For example, the researcher can take from this digitized book an entire lesson on how to label transcriptions to be submitted to the CHILDES database, and on the differences between video vs. audio transcriptions of language. This was wonderful – a great resource to someone who might not have access to the book.
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Very easy on Google Books. This is volume 1 of a third edition. The first two editions are included in Google Books, but are not available for preview.
MARC record available?	NA
Can the book be searched through an OPAC? If so, which one?	No
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	NA

SCAN SEARCH

search paradigm settings (example: every 10 double-pages)	scrolled through entire book

Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
3-4. 10-11, 17-18, 24-25, 31-32, 38-39, 45-46, all pages after 50 every other page in this book is of impeccable quality				

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	Comment (Is OCR search permitted): Y / N
Kanji script	successful detection, on two pages.
baby talk	successful recognition
zat	(a phonetic transcription provided on page 50 – detected successfully)
morphology	successful recognition
unintelligible speech	successful recognition
neologism	successful recognition
babbling	successful recognition
transcription	successful recognition
CHAT	successful recognition
@	(used as a phonetic transcription device in this book) successfully detected each time

IMAGE ASSESSMENT

If the original text includes graphics/illustrations, are they included in the electronic version?	No images in this book
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	
Is the scan on a gray-scale or color scale?	
Quality – does it look like the real thing?	
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	

Other notes

GENERAL OBSERVATIONS

Comment on ease of discovery in integrated interface (OPAC or better)	Very easy in Google Books.
Overall quality of scan of this book	Excellent, besides the fact that it is not a full preview option.
Ease of access to full text	To the available pages, very easy
Ability to copy, quote, cite?	No
Ability to search?	Yes
Ability to download?	No
Ability to print?	No
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	100% perfect
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Absolutely
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes
Issues with this book that would affect usefulness for scholars	None, besides that it is not fully available for view
Can this book be shared with other scholars/students?	Yes, by emailing the URL
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	NA
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	NA
Other notes	

Book URL:
http://books.google.com/books?id=6vMzhT1I6VsC&printsec=frontcover&dq=CHILDES&lr=&sig=ACfU3U0nmSWZhg-sNm90q0U2J9A_qQxLlg#PPA16,M1

Date accessed: July 27, 2008

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt **Date of investigation**: Aug 2, 2008

Book title: The sounds and inflections of the Greek dialects Ionic
Author: Herbert Weir Smyth
Year of Publication: 1894
Place of Publication: Clarendon Press
ISBN number?: na
Content/Disciplinary area: linguistics, Greek, descriptive grammar, dialect description

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	Provided by University of Michigan
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	very easily accessible in Google Books. Most metadata provided by Google on this book is missing: "Published by , 1894" No other editions
MARC record available?	NA
Can the book be searched through an OPAC? If so, which one?	No
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	Not provided by Google

SCAN SEARCH

search paradigm settings (example: every 10 double-pages) scrolled through entire book

Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
	No anomalies in this book! Overall great scan quality, pagination ordering			

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	**Comment**
	(Is OCR search permitted): Y / N
Aiolic	successful detection – also detected in small footnotes
Demeter	successful detection, but not in the footnote on page 623.
□□□□□□□□	unsuccessful detection (page 613)

□□□□□□	unsuccessful detection (page 613)
□□□□□□	unsuccessful detection (page 315)
□□□□□□□□	unsuccessful detection (page 315)
□□□□□	unsuccessful detection (page 268)
□	unsuccessful detection
□	unsuccessful detection
□	unsuccessful detection

IMAGE ASSESSMENT

If the original text includes graphics/illustrations, are they included in the electronic version?	No images in this text
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	
Is the scan on a gray-scale or color scale?	
Quality – does it look like the real thing?	
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	
Other notes	

GENERAL OBSERVATIONS

Comment on ease of discovery in integrated interface (OPAC or better)	Easily accessible in Google Books
Overall quality of scan of this book	Excellent – one of the best I've seen

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Ease of access to full text	100%
Ability to copy, quote, cite?	Yes, if you download document
Ability to search?	Yes – but NOT Greek characters, a major hindrance to any scholar studying Greek
Ability to download?	Yes
Ability to print?	Yes
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	Absolutely no recognition of Greek characters
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Yes - absolutely
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes
Issues with this book that would affect usefulness for scholars	NO OCR recognition of ANY Greek characters
Can this book be shared with other scholars/students?	Yes, by emailing URL
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Must fix OCR problem
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	NA
Other notes	

Book URL: http://books.google.com/books?id=X7oCAAAAMAAJ&pg=PA82&dq=Dionysios+Thrax&lr=&as_brr=1#PPA247,M1
Date accessed: Aug 2

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt	**Date of investigation**: August 21, 2008

Book title: A Grammar and Vocabulary of the Susoo Language to which are added the names of some of the susoo towns, near the banks of the rio pongas; a small catalogue of Arabic books, and a list of the names of some of the learners of the mandinga and foulah countries, with whom an useful correspondence could be opened up in the Arabic language.
Author: John Kemp
Year of Publication: 1803
Place of Publication: J. Ritchie, Blackfriars Wynd, London
ISBN number?: NA
Content/Disciplinary area: the Susoo language

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	Original provided by New York Public Library
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Easily accessible on Google Books – amount of metadata provided by Google Books is limited however
MARC record available?	No
Can the book be searched through an OPAC? If so, which one?	No
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	Metadata given is limited – no list of book chapters, for example. In the Table of Contents, the second, but not the first page, is hyperlinked (xliv)

SCAN SEARCH

search paradigm settings (example: every 10 double-pages)				
Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	**Comment**
	(Is OCR search permitted): Y / N

Sufoos	successful detection
Em fata fa gei	phrase on page 26, among others – successful detection
Kèmèë fhulifiring	Unsuccessful detection (on page 42)
Kèmèë	unsuccessful detection
fi*h*e	unsuccessful detection
Kôring	unsuccessful detection (on page 93)
Mungë	unsuccessful detection
shubenieng	successful detection
fûke	unsuccessful detection (on page 54)
ninkgeë	unsuccessful detection (also on page 54)

IMAGE ASSESSMENT	
If the original text includes graphics/illustrations, are they included in the electronic version?	No images in this text
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	
Is the scan on a gray-scale or color scale?	
Quality – does it look like the real thing?	
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	
Other notes	

GENERAL OBSERVATIONS

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Comment on ease of discovery in integrated interface (OPAC or better)	Easily accessible in Google Books
Overall quality of scan of this book	Excellent – good quality
Ease of access to full text	Yes
Ability to copy, quote, cite?	Yes
Ability to search?	Yes
Ability to download?	Yes
Ability to print?	Yes
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	Poor- most of the letters in the Susu language not detected
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Yes
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes
Issues with this book that would affect usefulness for scholars	No electronic searching of specific phonemes possible – this makes OCR searching nearly impossible with the Susu language
Can this book be shared with other scholars/students?	Yes, by emailing the URL
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	NA
Other notes	

Book URL: http://books.google.com/books?id=G_MSAAAAIAAJ&printsec=frontcover&dq=the+arabic+language&lr=#PPA54,M1

Date accessed: August 21

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt	**Date of investigation**: August 20, 2008

Book title: A Grammar of the Arabic Language
Author: E. H. Palmer
Year of Publication: 1874
Place of Publication: W. H. Allen & Co.
ISBN number?: no
Content/Disciplinary area: Arabic language, grammar

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	Original provided by Harvard University
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Easily accessible in Google Books. Lists some metadata in 'about this book' section
MARC record available?	NA
Can the book be searched through an OPAC? If so, which one?	No
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	Metadata provided is sufficient- does not, for example, provide chapters but rather lists some content found in the book. Table of contents in the book is hyperlinked.

SCAN SEARCH

search paradigm settings (example: every 10 double-pages)	searched through entire book as to not miss anything

Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
138		bottom left hand corner is whited out		
221	this page was moved during scan – completely illegible		see figure 2	
no other anomalies in this book				

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Word	Comment
(Is OCR search permitted): Y / N	
□	unsuccessful detection
□	unsuccessful detection
□□	unsuccessful detection
aortist	successful
Hollow verbs	successful
□	unsuccessful detection
□	unsuccessful detection
□	unsuccessful detection
□	unsuccessful detection
□	unsuccessful detection

IMAGE ASSESSMENT	
If the original text includes graphics/illustrations, are they included in the electronic version?	No images in this text, but there are several graphs and charts with Arabic phonemes
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	
Is the scan on a gray-scale or color scale?	
Quality – does it look like the real thing?	
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	
Other notes	

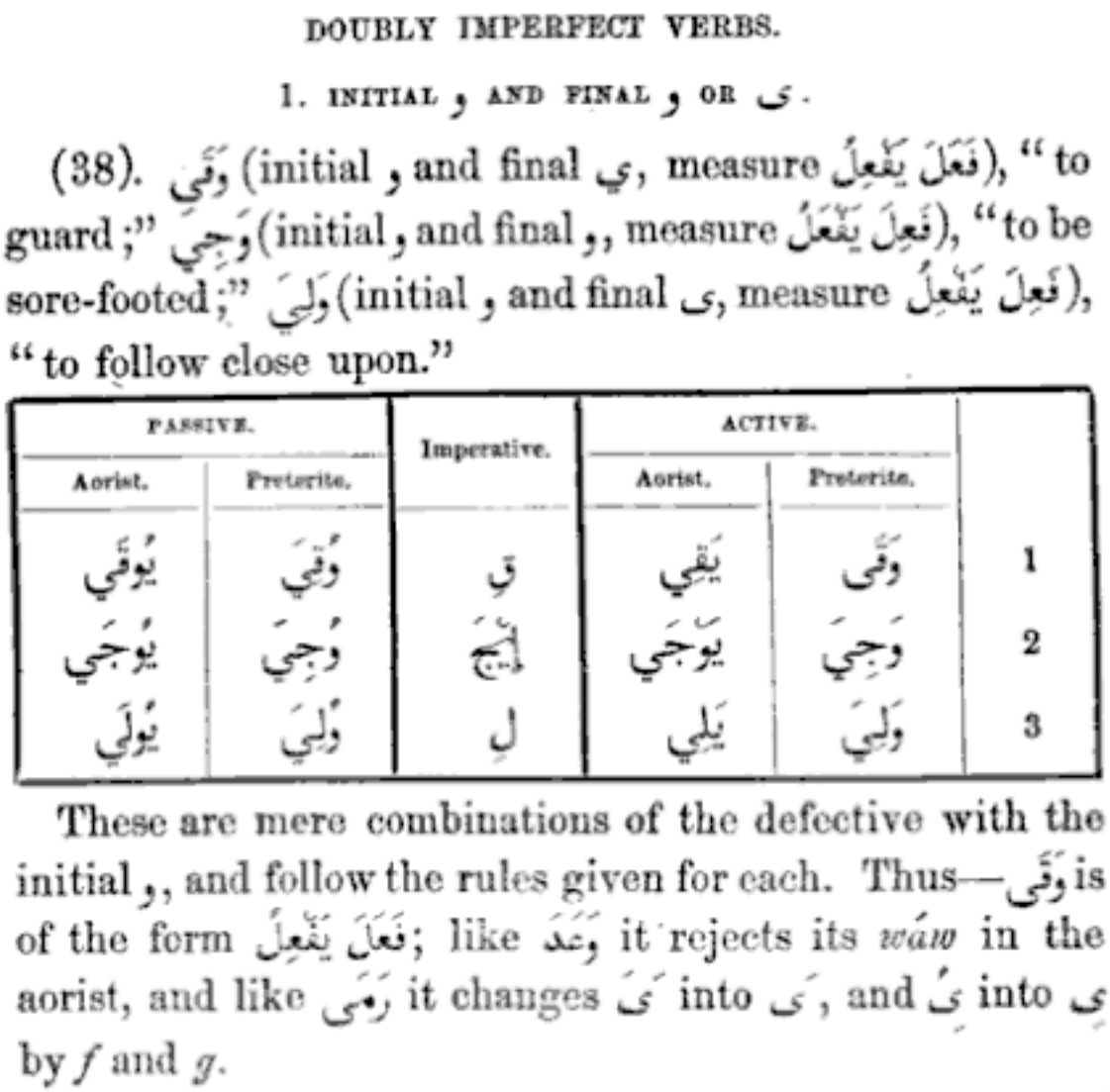
CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

GENERAL OBSERVATIONS	
Comment on ease of discovery in integrated interface (OPAC or better)	Easily accessible in Google Books
Overall quality of scan of this book	Good- the very tips of the scanner's fingerprints are visible in many of the pages
Ease of access to full text	Yes
Ability to copy, quote, cite?	Yes
Ability to search?	Yes
Ability to download?	Yes
Ability to print?	Yes
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	No detection of Arabic letters at all
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Yes
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes
Issues with this book that would affect usefulness for scholars	No electronic searches of Arabic are possible in this book
Can this book be shared with other scholars/students?	Yes, by emailing the URL to students
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	NA
Other notes	Index in back of book has all pages hyperlinked

Book URL: http://books.google.com/books?id=gdHjvHF-E2kC&printsec=frontcover&dq=the+arabic+language&lr=#PPA405,M1

Date accessed: August 20

Figure 1 example of Arabic in this book



DOUBLY IMPERFECT VERBS.

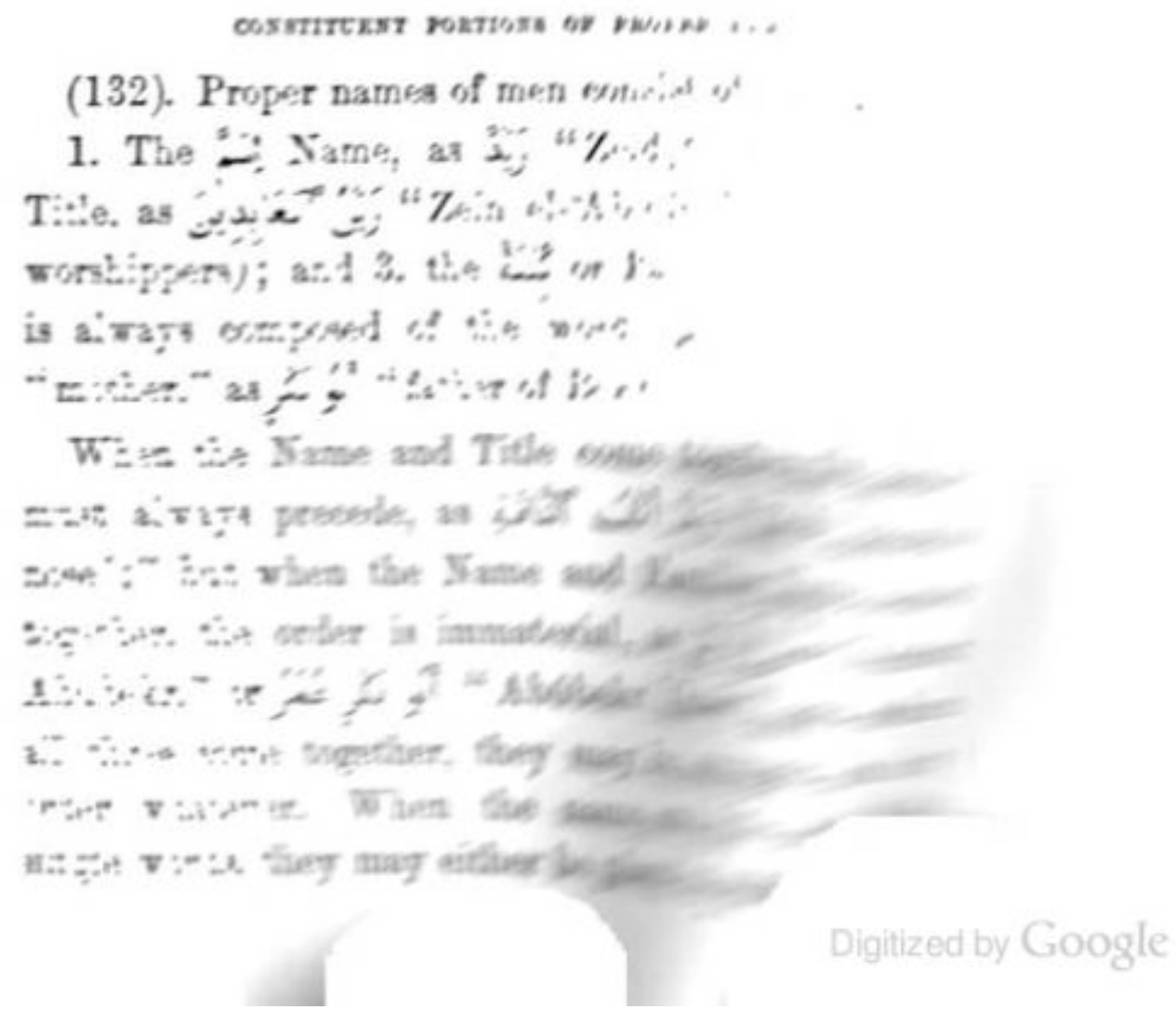
1. INITIAL و AND FINAL و OR ى.

(38). وَقَى (initial و and final ي, measure فَعَلَ يَفْعِلُ), "to guard;" وَجِيَ (initial و and final و, measure فَعِلَ يَفْعَلُ), "to be sore-footed;" وَلِيَ (initial و and final ى, measure فَعِلَ يَفْعِلُ), "to follow close upon."

PASSIVE.		Imperative.	ACTIVE.		
Aorist.	Preterite.		Aorist.	Preterite.	
يُوقَى	وُقِيَ	قِ	يَقِي	وَقَى	1
يُوجَى	وُجِيَ	جِيَ	يَوْجَى	وَجِيَ	2
يُولَى	وُلِيَ	لِ	يَلِي	وَلِيَ	3

These are mere combinations of the defective with the initial و, and follow the rules given for each. Thus—وَقَى is of the form فَعَلَ يَفْعِلُ; like وَعَدَ it rejects its *wáw* in the aorist, and like رَمَى it changes يَ into ى, and يُ into ي by *f* and *g*.

Figure 2
Page 221, scan moved



CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt **Date of investigation**: August 19, 2008

Book title: A Practical Grammar of the Arabic Language
Author: Faris el-Shidiac, 2nd edition by Henry G. Williams
Year of Publication: 1866
Place of Publication: Bernard Quaritch, London
ISBN number?: na
Content/Disciplinary area: Arabic language

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	original provided by Harvard College.
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Easily accessible on Google Books. No links to the first edition or to other words by the author.
MARC record available?	NA
Can the book be searched through an OPAC? If so, which one?	No
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	Not very detailed – lists 'sections' for example, as opposed to exact chapters of book

SCAN SEARCH

search paradigm settings (example: every 10 double-pages) searched through entire book

Page(s):	readability/ quality of scan	anomalies?	other?	other?
2, 3	an apparently hand-written alphabet of Arabic is provided on these pages – none of the letters are searchable in OCR scan			
very last page- a scan of the back binder, shows scanner's finger	no other anomalies in this book!			

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	Comment (Is OCR search permitted): Y / N
□	not detected
□	not detected
□	not detected
□□□	not detected
gutteral	not detected- this word is present on page 4
conversation	successfully detected
alif	successfully detected
preterite	successfully detected
imperative	successfully detected
conjunctions	successfully detected

IMAGE ASSESSMENT

If the original text includes graphics/illustrations, are they included in the electronic version?	No images in this book
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	
Is the scan on a gray-scale or color scale?	
Quality – does it look like the real thing?	
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	
Other notes	

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

GENERAL OBSERVATIONS

Comment on ease of discovery in integrated interface (OPAC or better)	Easily accessible in Google Books, not on any OPAC
Overall quality of scan of this book	excellent
Ease of access to full text	Yes
Ability to copy, quote, cite?	Yes
Ability to search?	Yes – but no detection of any Arabic letters
Ability to download?	Yes
Ability to print?	Yes
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	Poor
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Yes
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes
Issues with this book that would affect usefulness for scholars	Electronic detection of Arabic, and even some English words, is not possible. This would prevent scholars from doing research in the Arabic language.
Can this book be shared with other scholars/students?	Yes, by emailing the URL
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	NA
Other notes	

Book URL: http://books.google.com/books?id=IilZghQnp4sC&printsec=frontcover&dq=the+arabic+language#PPA44,M1
Date accessed: August 19, 2008

Figure 1 Example of Arabic

THE Arabs have but one article, اَلْ (*al*), which is definite, and is prefixed either to the singular or plural; as, اَلْكِتَاب *alkitāb*, "the book;" اَلْكُتُب *alkotob*, "the books." When the article is followed by any of these letters, viz. ن, ظ, ط, ض, ص, ش, س, ز, ر, ذ, د, ث, ت, the sound of the ل is dropped; or, rather, it is assimilated to that of the succeeding letter, which assumes the تَشْدِيد *tashdeed*; as اَلرَّحِيم *arraheem*, "the merciful" (*i.e.* God); اَلسَّمَوَات *assamawāt*, "the heavens;" فِقْهُ الدِّين *fik'hoddeen*, "knowledge of the religion."

N.B. The word to which the article is annexed does not admit اَلتَّنْوِين the *tanween*.

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt **Date of investigation**: August 9, 2008

Book title: Ancient India
Author: Edward James Rapson
Year of Publication: 1922
Place of Publication: Macmillan
ISBN number?: na
Content/Disciplinary area: India; this book discusses potential for proto-Persian language, which was my main interest

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	Original provided by Harvard University.
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	This book is in a volume series (6 total), however, Google metadata does not provide information on this. Table of Contents, for example, is not provided as metadata.
MARC record available?	Na
Can the book be searched through an OPAC? If so, which one?	No
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	Metadata provided is limited. However, actual Table of contents within book is hyperlinked – a great resource for scholars. Also, back of the book index has several of its page numbers hyperlinked.

SCAN SEARCH

search paradigm settings (example: every 10 double-pages) Searched through entire book

Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
4-5; 25-26; 345-346; 374-375 (actually, pages in between these, as they're not numbered in the book); 514-515;	On these pages, a map of India is provided (two page layout original format). It's division into separate pages of a papyrus roll makes it a bit awkward. Scan is not top quality			
58, 278 (as an example)		sometimes scanner's fingertips show up in the scan		
291		preceding page is caught in the scan		
463	some unique symbols are seemingly hand-drawn on this page			

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	Comment
	(Is OCR search permitted): Y / N
Dravidian	successful detection – with yellow highlighting
Toalas	successful detection (on page 41 this word is in much smaller and slightly blurred font- it was detected perfectly however)
Indo-Aryans	successful detection
Austric	successful detection
Munda	successful detection
□□□□□□□□, □□□□□	□□□ successful – both of these Greek words are on page 453, OCR search did not detect them
Persian	successful detection
Bactria	successful detection
Aramaic	successful detection
Proto-Iranian	successful detection

IMAGE ASSESSMENT

If the original text includes graphics/illustrations, are they included in the electronic version?	In the back of the book, there are several pictures of Indian artifacts, coins and architecture.
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	Not sure, as these are actual pictures
Is the scan on a gray-scale or color scale?	grey
Quality – does it look like the real thing?	yes
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	No illustrations, besides the maps listed above
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	

Other notes

GENERAL OBSERVATIONS

Comment on ease of discovery in integrated interface (OPAC or better)	Easily accessible in Google books
Overall quality of scan of this book	Excellent
Ease of access to full text	Yes
Ability to copy, quote, cite?	Yes, however, Greek font and perhaps other alphabets are not OCR detectable
Ability to search?	Yes
Ability to download?	Yes
Ability to print?	Yes
Scanned book allow for consulting of contents?	Yes, besides Greek font
How successful is OCR recognition?	Perfect for English alphabet, English language-based searches.
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Yes
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes
Issues with this book that would affect usefulness for scholars	No searchability of Greek or other languages in this book, which is a limitation
Can this book be shared with other scholars/students?	Yes, by emailing the URL
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	No – however other volumes in this series are not provided by Google Books
Other notes	

Book URL: http://books.google.com/books?id=nn0TAAAAYAAJ&pg=RA1-PA320&dq=proto+persian&lr=&as_brr=1#PPR16,M1

Date accessed: Aug 9 2008

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt	**Date of investigation**: August 23, 2008

Book title: Archivo Boliviano collecciones de documentos relativos a la historia de Bolivia, durante la época colonial, con un catálogo de obras impresas y de manuscritos, que tartan de esa parte de la América Meridional.
Author: Vicente de Ballivian y Róxas
Year of Publication: 1872
Place of Publication: France, A. Franck (f. Vieweg)
ISBN number?: NA
Content/Disciplinary area: historic documents of Bolivia

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	Google does not list who the original book is provided by. Scanned pages at end indicate it was Stanford University
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Easily accessible in Google Books
MARC record available?	No
Can the book be searched through an OPAC? If so, which one?	No
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	Book metadata incorrectly lists author as Vicente de Ballivián y Rojas. His original name is spelled Róxas. It's important to maintain original spelling in old Spanish texts. Google metadata also lists two other authors - they are not the actually publisher of the collection as Róxas is.

SCAN SEARCH

search paradigm settings (example: every 10 double-pages)		scanned through entire book		
Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
intro		scanner's fingers present in this page (Figure 1)		
486		scanner's fingers present in this page (Figure 2)		
502	blurriness in top right corner of page			

WORD SEARCHES DONE WITH THIS BOOK

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)	
Word	**Comment**
(Is OCR search permitted): Y / N	
la conquista	successful detection
Tipuani	successful detection
Arsans	successful detection
Cantumarca	successful detection
Potosi, Potosí	successful detection (accentuation does make a difference)
estremeños	successful detection
indios	successful detection
Vicuñas	successful detection
admirarnos	successful detection (small text in a footnote)
andaluz	successful detection

IMAGE ASSESSMENT	
If the original text includes graphics/illustrations, are they included in the electronic version?	No images in this text
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	
Is the scan on a gray-scale or color scale?	
Quality – does it look like the real thing?	
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	
Other notes	

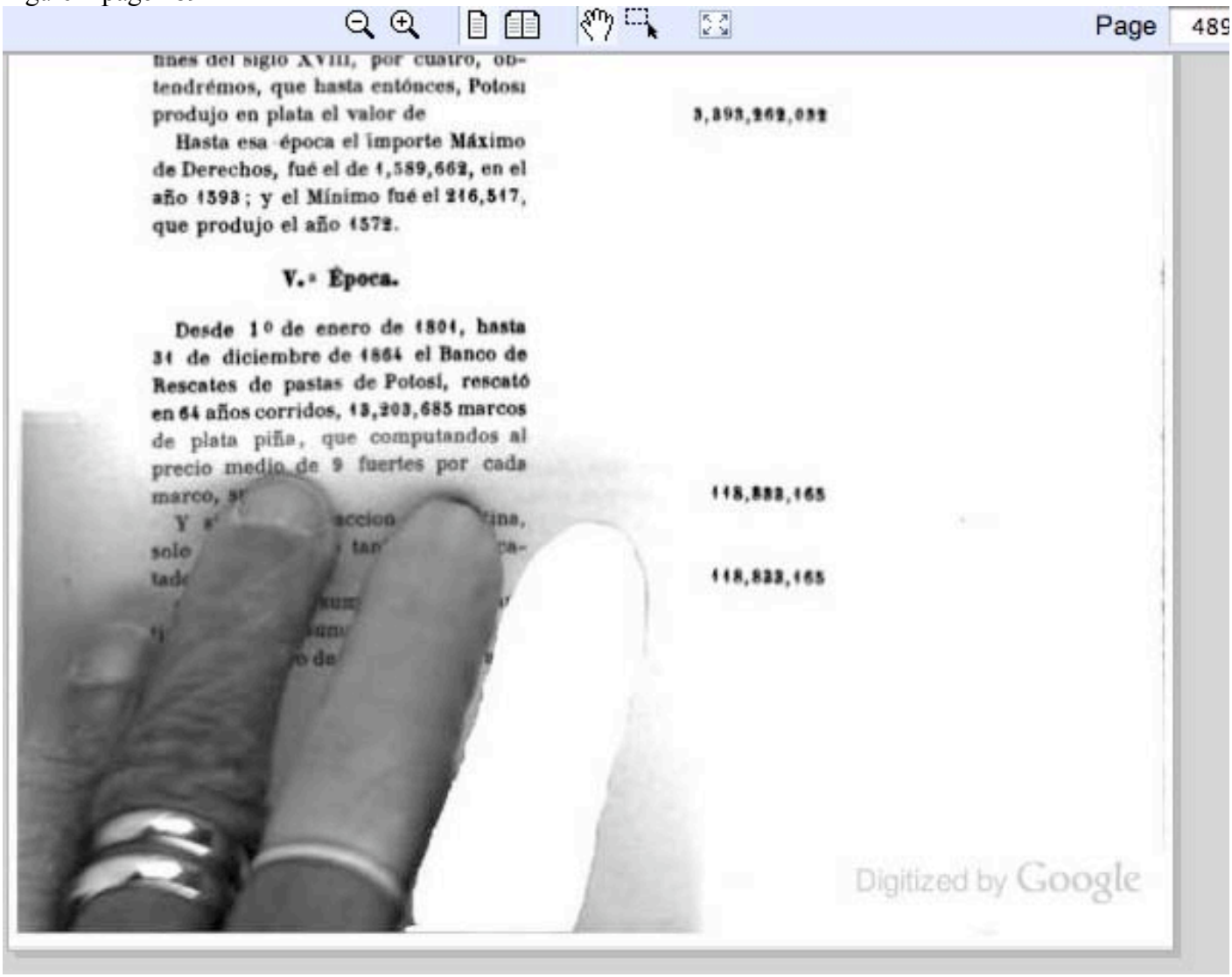
CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

GENERAL OBSERVATIONS	
Comment on ease of discovery in integrated interface (OPAC or better)	Easily accessible in Google
Overall quality of scan of this book	Good
Ease of access to full text	Yes
Ability to copy, quote, cite?	Yes
Ability to search?	Yes
Ability to download?	Yes
Ability to print?	Yes
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	seemingly perfect
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Yes
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes
Issues with this book that would affect usefulness for scholars	None – this book overall has no anomalies
Can this book be shared with other scholars/students?	yes, but emailing the URL
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	NA
Other notes	

Book URL:
http://books.google.com/books?id=lSZK01ZIjiUC&printsec=frontcover&dq=potosi&lr=lang_es&as_brr=1&source=gbs_similarbooks_r&cad=3_2#PPT1,M1
Date accessed: August 23, 2008

Figure 1
Introduction to book – scanner's fingers present

Figure 2 page 489



CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt	**Date of investigation**: August 22, 2008

Book title: La literatura argentina desde la conquista hasta nuestros días: Seguida de un estudio sobre la literatura de los demás países Hispano-americanos
Author: Felipe Martinez
Year of Publication: 1905
Place of Publication: Imprenta M. Biedma e Hijo, Bolivar 535
ISBN number?: NA
Content/Disciplinary area: Argentina, literature

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	Original provided by University of Michigan
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Easily accessible on Google Books, metadata provided is sufficient.
MARC record available?	No
Can the book be searched through an OPAC? If so, which one?	No
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	Not too much detail is provided – minimal information, for example, is given on book "sections"

SCAN SEARCH

search paradigm settings (example: every 10 double-pages)	scrolled through entire book

Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
	no major anomalies in this book			

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	**Comment**
	(Is OCR search permitted): Y / N

composición	successful detection
Cicerón	successful detection
líricamente	successful detection
Los Consuelos	successful detection
bálsamo	successful detection
ilumina	successful detection
ensueño	successful detection
penumbra	successful detection
carnaval	successful detection
Buenos Aires	successful detection

IMAGE ASSESSMENT	
If the original text includes graphics/illustrations, are they included in the electronic version?	No images in this text
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	
Is the scan on a gray-scale or color scale?	
Quality – does it look like the real thing?	
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	
Other notes	

GENERAL OBSERVATIONS

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Comment on ease of discovery in integrated interface (OPAC or better)	Easily accessible in Google Books
Overall quality of scan of this book	good
Ease of access to full text	Yes
Ability to copy, quote, cite?	Yes
Ability to search?	Yes
Ability to download?	Yes
Ability to print?	Yes
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	perfect
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Yes
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes
Issues with this book that would affect usefulness for scholars	None – this book overall has no issues – a good quality scan
Can this book be shared with other scholars/students?	Yes, by emailing the URL
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	NA
Other notes	

Book URL: http://books.google.com/books?id=7rUeAAAAMAAJ&pg=PA226&dq=argentina,+literatura,+gobierno+date:1700-1980&lr=lang_es&as_brr=0#PPA163,M1

Date accessed: August 22, 2008

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt **Date of investigation**: 6/16/08

June 16, 2008

Search record

Today I am trying to find complete books available online that were written before 1900. I'm looking for original texts in original languages. This proved to be quite daunting – it seems that Google mostly offers English versions that are recent, as opposed to originals. I did the following searches until I found a book to examine:

1. Marco Polo
2. Travels of Dean Mahomet (1794, M.H. Fisher) – a limited edition was available, but not original.
3. Tales of Moonlight and Rain (1776, Akinari Ueda)
4. The Dancing Girl (1790, Hasan Shah). This book was originally written in Farsi. Google Book's only offers a limited-view of the English translated version by Qurratulain Hyder, 1993. Even after designating language results to Persian, still no original available.
4. The Anatomy of Melancholy, by Robert Burton (1621). A hit! This is the book I'll evaluate today.

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Book title: Anatomy of Melancholy v. 1
Author: Robert Burton
Year of Publication: originally 1621. The one I am researching is an edition in three volumes, by W. J. Widdleton Publisher, 1875 (this book is part of Lane Medical Library, Stanford)
Place of Publication: New York
ISBN number?:
Content/Disciplinary area: Medical field, history (one of first attempts to describe depression and ailments)

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	Digitized version excellent. Original volumes from 1621 – not sure where those are or their condition/fate. The 1875 Widdleton version is the version online.
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Very easy, however, there are ten versions online with Google Book. Most have no preview available, two versions have limited preview. Two versions have full view – these are the versions I am exploring. The first, titled volume 1, makes up this examination. The second, ALSO TITLED volume 1 by Google Books, is actually volume 3 (mistitled). Volume 2 appears to be missing.
MARC record available?	No
Can the book be searched through an OPAC? If so, which one?	Yes! Georgetown's library site has 11 different MARK records for this book. With the first result - An 'inside book' preview for the titled available at Qatar Stacks is provided by Syndetics and includes table of contents, summary and cover image. For the other versions available in GEORGE (Georgetown's Library Catalog), I really had to try each link to see which ones led to online versions available. This is not immediately indicated with a GEORGE result listing. The researcher must click on all search results and will then learn which have available online versions. For example, GEORGE results 3, 4, 5, 8, and 9 (out of the 11 search results) have links to online versions. (It might be more efficient if online options were shown along with the results hit.) All online versions are provided by EEBO – Early English Books Online. It seems EEBO has several different editions of Burton's book, corresponding with each GEORGE result. I'm attaching a pdf file of the 1676 version available by EEBO. Within EEBO, I can change view preference, choosing either record, thumbnails, illustration, or document images. (The durable URL for this visible document is http://0-gateway.proquest.com.library.lausys.georgetown.edu/.) I can also download as either TIFF or PDF file, but I can ONLY download each visible page at a time – not the whole document as one file. While it's wonderful to have a digitized copy of the original 1961, 1676 versions, etc., I unfortunately cannot do any searches in it with EEBO (let alone download entire document at one). I'll thus stick to the Google doc for now.
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	NA – nice introduction by publishers on author. Beautiful inner page of illustrations.

SCAN SEARCH

search paradigm settings (example: every 10 double-pages)	Every 10 pages. I scrolled through pages as to not miss anything. The pages of this book correspond with Google's page numbering – 1:1, 2:2, …60:60, etc. Very nice.

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Page(s):	readability/ quality of scan	anomalies?	other?	other?
1	no text			
10	excellent	none		
20, 30	excellent	discoloration in bottom left		
40, 50	excellent	none		
59, 59 (second)	First scan of page 59 is excellent. There's then a page 59 repeat, with a scan of page being turned!	complete blurriness - illegible	(see attached picture)	Google could just eliminate the second page 59. Interestingly, the page orderings are not affected- page 59 just happens twice.
61	skips page 60 – straight to 61, and there's a human hand on the scan	bottom half illegible	(see attached picture)	
60-61	excellent	none	Page 60 and 61 are then repeated, in excellent condition. Google book page numbers also correspond perfect with scanned pages.	Good could just eliminate the extra pages with anomalies.
70	excellent	some discoloration in bottom left corner	does not impede reading of text	
80, 90	excellent	none		
93	impossible- page was moved	completely illegible, scanner unfortunately never rescanned this page	Page 93 therefore omitted from this scan – entire page is one complete blur	THIS PAGE MISSING FROM SCANNED DOCUMENT
100, 110,	excellent	none		
120, 130, 140	excellent		black outline of scanner's fingers is present on lower bottom left corner of page	
150, 160	excellent	some discoloration in bottom left corner		
170, 180, 190, 200, 210, 220, 230, 240 = 340, 350, 360, 370, 380, 390 = 420, 430 = 490	excellent	none		[= means I jumped to next page sequence]

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

(Is OCR search permitted): Y / N

word	comment
saxoniâ	I saw this word in the text with symbol over the a, and wanted to see if a search with this symbol would yield results. Saxoniâ does NOT, but the browser did find saxonia (no symbol).
body	highlighted all finds in yellow. Success.
debet	This was in a footnight – successful.
aliena	Successful, however, this word is in a footnote on page 52, and it was not recognized. (In fact, in searching for 'debet' above, which is proximal to the word aliena in the footnote, pg. 52, the sentence provided in the results with 'debet' shows aliena as 'aliens'! It mistakes the a for an a.
depression	highlighted in yellow, successful.
cure	highlighted in yellow, successful. I noticed that on page 24 (text is very small), the sentence provided in the 'cure' result hit shows jail as 'j dl'
woman	highlighted in yellow, successful.

God	highlighted in yellow, successful.
sacerdote	(A lot of Latin is in this book). On page 99, there is an italicized sacerdotes. Browser did not recognize 'sacerdote' from the 'sacerdotes.' Once I entered 'sacerdotes' (with an –s) did the search yield a successful result. I'd think that it would find the word even within others, regardless of plural morphology, etc.
organs	highlighted in yellow, successful.
□, □	There also is some Greek present in this book. I noticed a couple of words with pi and/or phi in them, attempted a search with these letters. Search unsuccessful, even though there are many instances of this symbol.

Note- some pages have words that curve, going along with the shape of the book's inner binder. I tried a few searches on these and it was successful.

GENERAL OBSERVATIONS

Comment on ease of discovery in integrated interface (OPAC or better)	Very easy – though both GEORGE (Georgetown's catalog) and Google have several different versions. The researcher must look at each one to see which editions have online editions. (and in Google's case, which one's allow viewing)
Overall quality of scan of this book	High quality – except for one missing page and several incidents of the scanner's fingers showing in the scan
Ease of access to full text	Very
Ability to copy, quote, cite?	Possibly – if you download the version from Google.
Ability to search?	Yes
Ability to download?	Yes!
Ability to print?	yes
Scanned book allow for consulting of contents?	yes
How successful is OCR recognition?	excellent
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Yes, except for the fingers that are scanned. There are numerous pages in this book where scanned fingers show.
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes
Issues with this book that would affect usefulness for scholars	None – besides one missing page
Can this book be shared with other scholars/students?	Yes, by emailing link
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	Volume 2 is missing from Google Books, Volume 3 is mistitled as Volume 1. (They therefore have two volumes titled Volume 1).
Other notes	In going from page to page, loading of images and text was slow at times. However, I loved that this book allows me to jump from page to page by entering page number. This text is not papyrus-like, as other google books are.

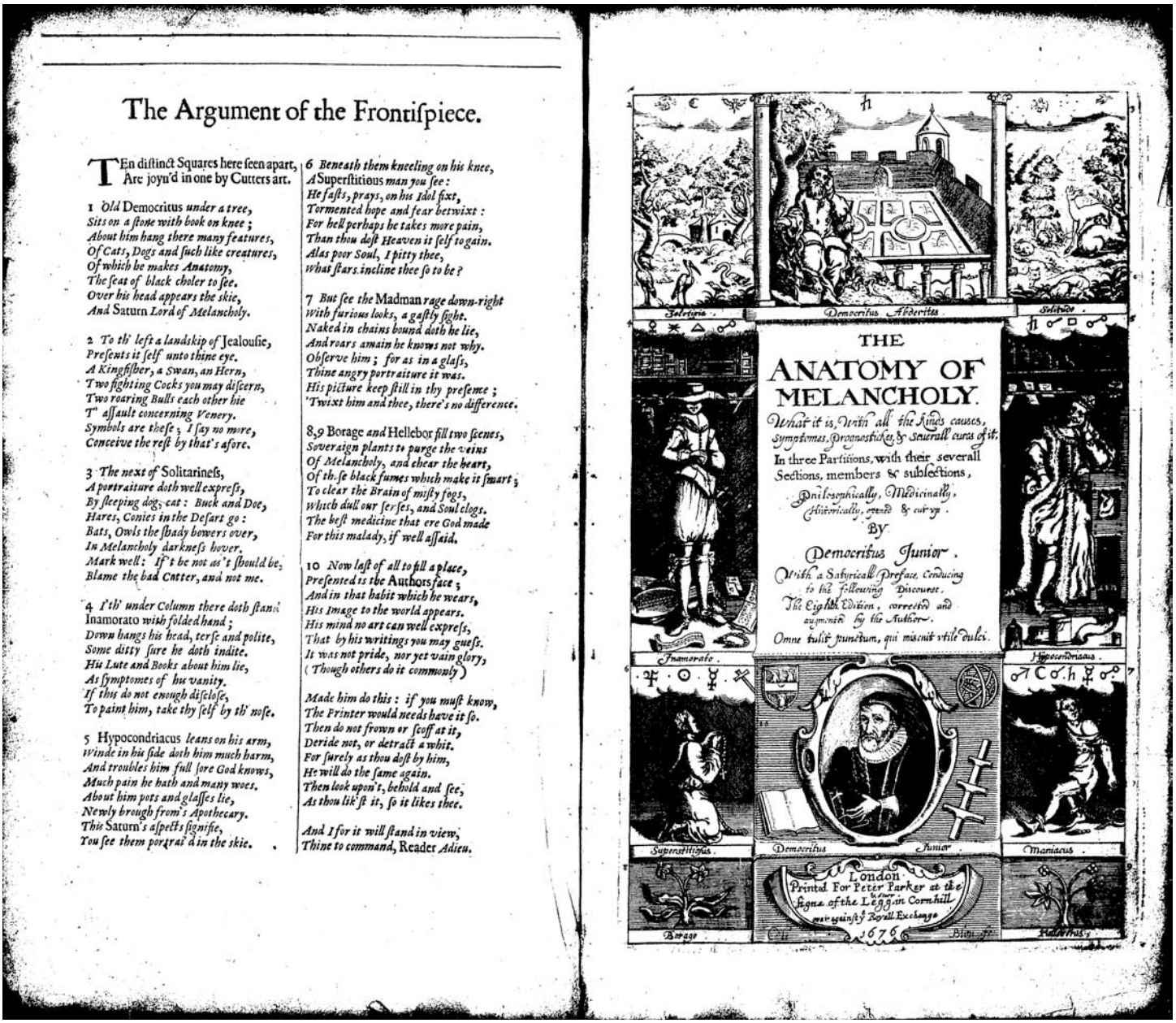
Book URL:
http://books.google.com/books?id=R0YJAAAAIAAJ&printsec=frontcover&dq=editions:0YP4Pr8J8AP6uDLt5yNC#PPA392,M1

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Date accessed: 6/16/08

Example of EEBO thumbnail view of 1676 version. Researchers can download each individual page as TIFF or PDF files, but not the entire document at once.

Example of downloaded PDF file of 1676 version, provided by EEBO. Researchers may connect to EEBO after doing searches in GEORGE's catalog and finding editions with online versions. Quality is wonderful – no OCR search capabilities, however.



The Argument of the Frontiſpiece.

TEn diſtinct Squares here ſeen apart,
Are joyn'd in one by Cutters art.

1 Old Democritus *under a tree,*
Sits on a ſtone with book on knee;
About him hang there many ſeatures,
Of Cats, Dogs and ſuch like creatures,
Of which he makes Anatomy,
The ſeat of black choler to ſee.
Over his head appears the skie,
And Saturn *Lord of Melancholy.*

2 *To th' left a landskip of* Jealouſie,
Preſents it ſelf unto thine eye.
A Kingfiſher, a Swan, an Hern,
Two fighting Cocks you may diſcern,
Two roaring Bulls each other hie
T' aſſault concerning Venery.
Symbols are theſe; I ſay no more,
Conceive the reſt by that's afore.

3 *The next of* Solitarineſs,
A portraiture doth well expreſs,
By ſleeping dog, cat: Buck and Doe,
Hares, Conies in the Deſart go:
Bats, Owls the ſhady bowers over,
In Melancholy darkneſs hover.
Mark well: If't be not as't ſhould be,
Blame the bad Cutter, and not me.

4 *I'th' under Column there doth ſtand*
Inamorato *with folded hand;*
Down hangs his head, terſe and polite,
Some ditty ſure he doth indite.
His Lute and Books about him lie,
As ſymptomes of his vanity.
If this do not enough diſcloſe,
To paint him, take thy ſelf by th' noſe.

5 Hypocondriacus *leans on his arm,*
Winde in his ſide doth him much harm,
And troubles him full ſore God knows,
Much pain he hath and many woes.
About him pots and glaſſes lie,
Newly brought from's Apothecary.
This Saturn's *aſpects ſignifie,*
You ſee them portrai'd in the skie.

6 *Beneath them kneeling on his knee,*
A Superſtitious *man you ſee:*
He faſts, prays, on his Idol fixt,
Tormented hope and fear betwixt:
For hell perhaps he takes more pain,
Than thou doſt Heaven it ſelf to gain.
Alas poor Soul, I pitty thee,
What ſtars incline thee ſo to be?

7 *But ſee the* Madman *rage down-right*
With furious looks, a gaſtly ſight.
Naked in chains bound doth he lie,
And roars amain he knows not why.
Obſerve him; for as in a glaſs,
Thine angry portraiture it was.
His picture keep ſtill in thy preſence;
'Twixt him and thee, there's no difference.

8,9 Borage *and* Hellebor *fill two ſcenes,*
Soveraign plants to purge the veins
Of Melancholy, and chear the heart,
Of th.ſe black fumes which make it ſmart;
To clear the Brain of miſty fogs,
Which dull our ſenſes, and Soul clogs.
The beſt medicine that ere God made
For this malady, if well aſſaid.

10 *Now laſt of all to fill a place,*
Preſented is the Authors *face;*
And in that habit which he wears,
His Image to the world appears.
His mind no art can well expreſs,
That by his writings you may gueſs.
It was not pride, nor yet vain glory,
(Though others do it commonly)

Made him do this: if you muſt know,
The Printer would needs have it ſo.
Then do not frown or ſcoff at it,
Deride not, or detract a whit.
For ſurely as thou doſt by him,
He will do the ſame again.
Then look upon't, behold and ſee,
As thou lik'ſt it, ſo it likes thee.

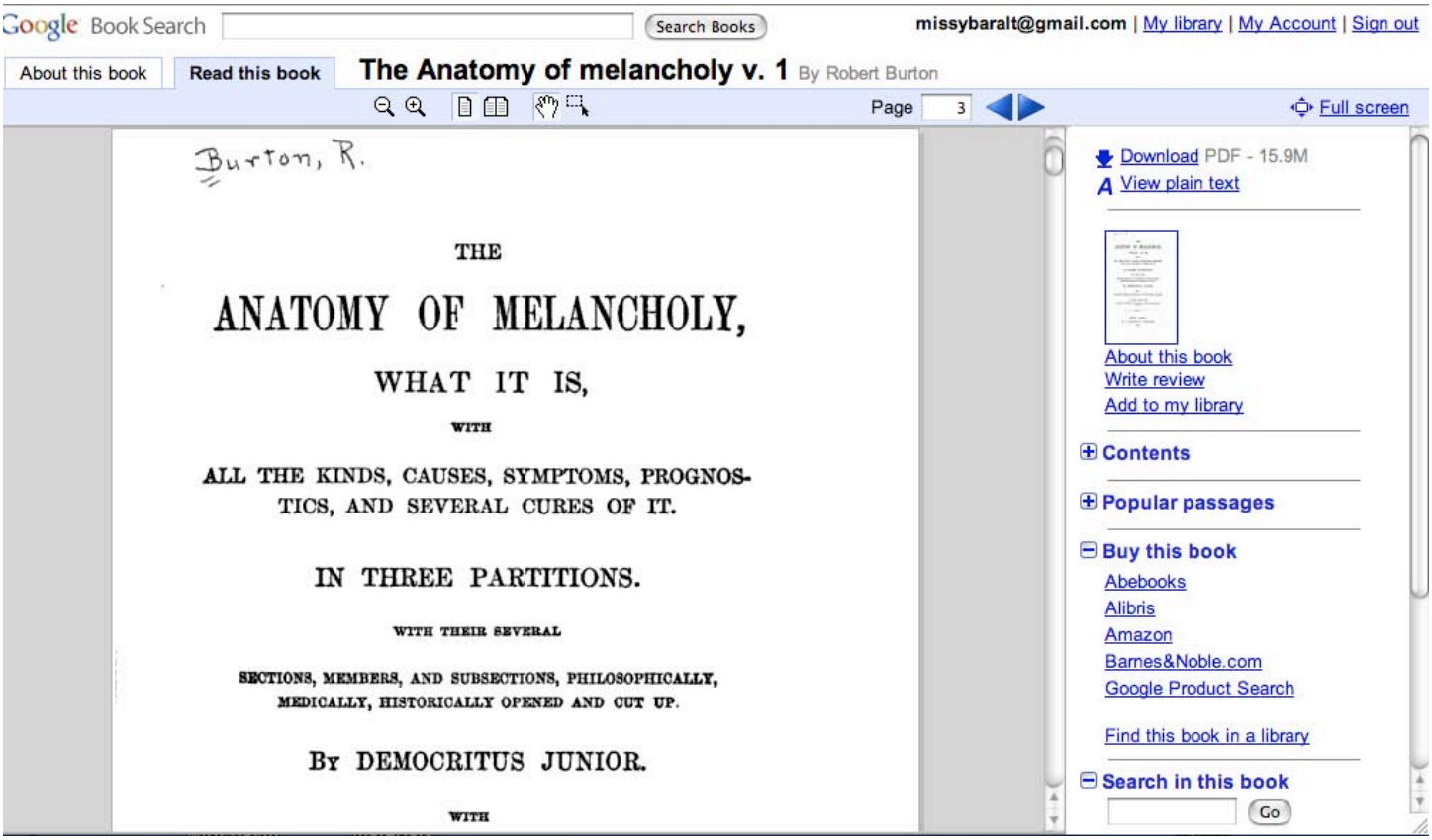
And I for it will ſtand in view,
Thine to command, Reader *Adieu.*

Democritus Junior.

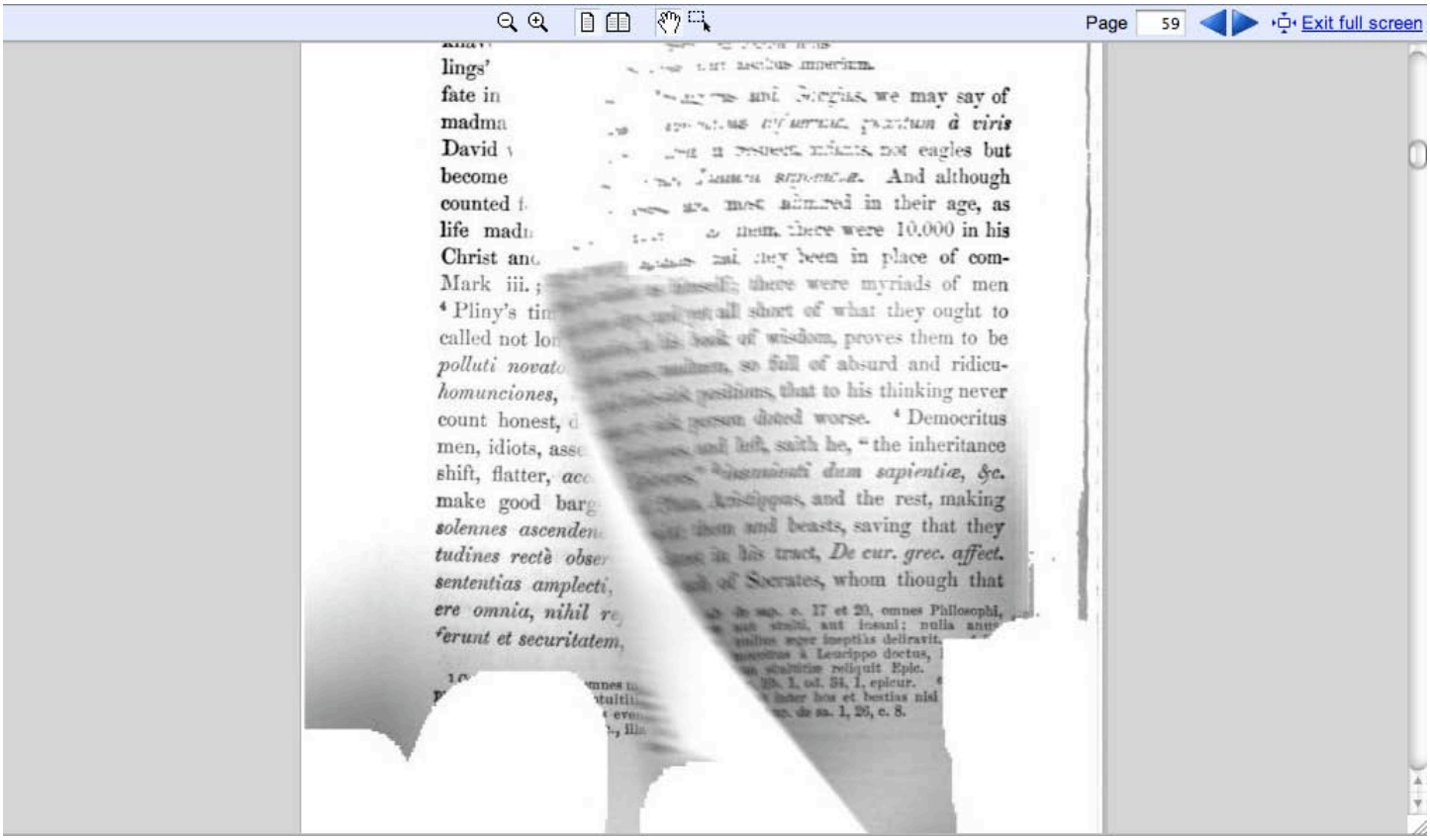
THE
ANATOMY OF
MELANCHOLY.
What it is, With all the Kinds causes, Symptomes, Prognostickes, & Severall cures of it.
In three Partitions, with their severall Sections, members & subsections,
Philosophically, Medicinally, Historically, opened & cut up.
By
Democritus Junior.
With a Satyricall Preface, Conducing to the following Discourse.
The Eighth Edition, corrected and augmented by the Author.
Omne tulit punctum, qui miscuit utile dulci.

London
Printed For Peter Parker at the Sign of the Leg & in Cornhill, neer unto the Royall Exchange.
1676

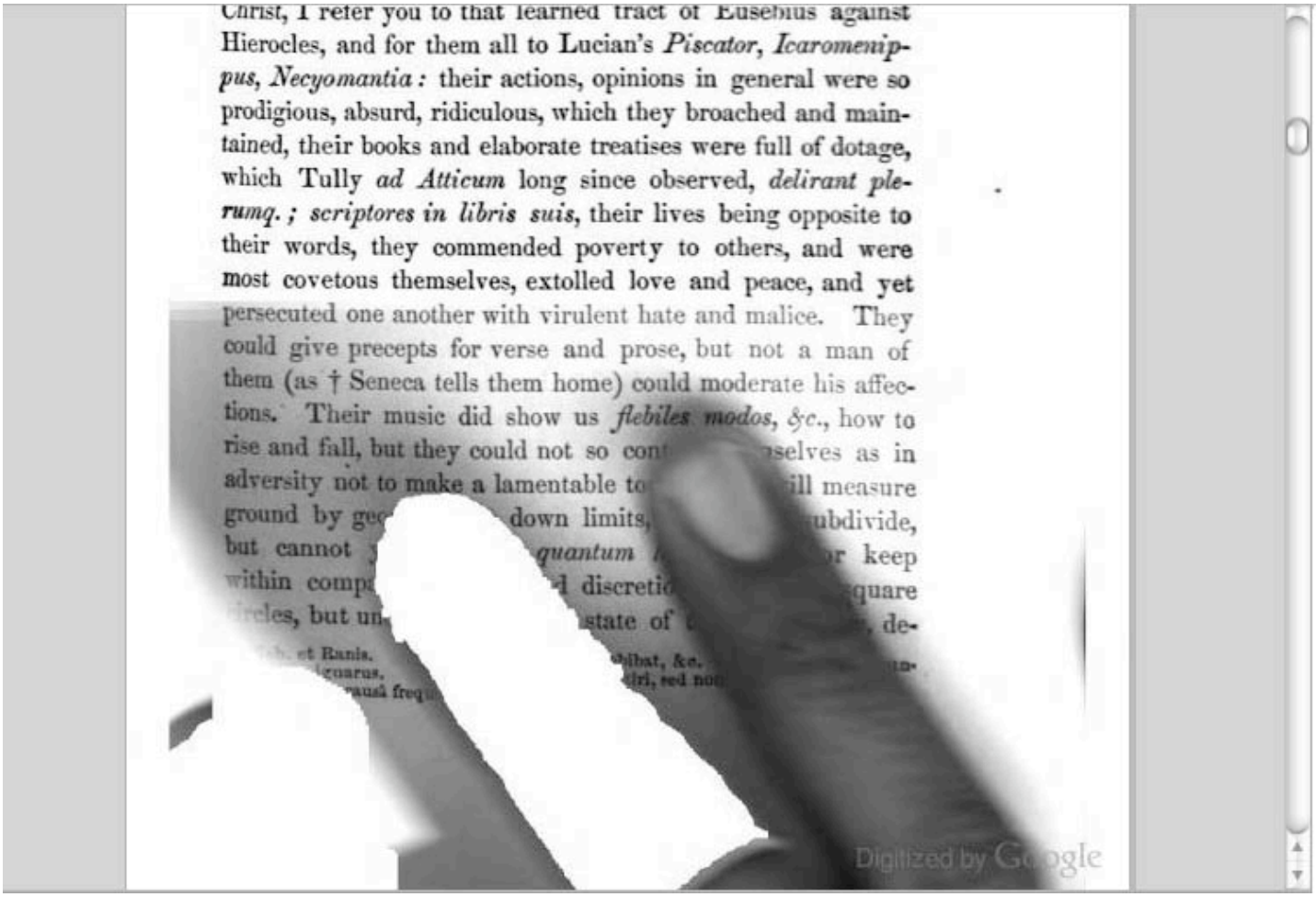
Google Book Image Example (this digitized version is an 1875 edition by Widdleton Publisher).



Page 59, scan of page turn



Page 61, scanner's hand in scan



Christ, I refer you to that learned tract of Eusebius against Hierocles, and for them all to Lucian's *Piscator, Icaromenippus, Necyomantia:* their actions, opinions in general were so prodigious, absurd, ridiculous, which they broached and maintained, their books and elaborate treatises were full of dotage, which Tully *ad Atticum* long since observed, *delirant plerumq.; scriptores in libris suis*, their lives being opposite to their words, they commended poverty to others, and were most covetous themselves, extolled love and peace, and yet persecuted one another with virulent hate and malice. They could give precepts for verse and prose, but not a man of them (as † Seneca tells them home) could moderate his affections. Their music did show us *flebiles modos, &c.*, how to rise and fall, but they could not so con[illegible]selves as in adversity not to make a lamentable to[illegible]ill measure ground by ge[illegible] down limits, [illegible]ubdivide, but cannot [illegible] *quantum* [illegible]r keep within comp[illegible]d discreti[illegible]quare [illegible]rcles, but un[illegible] state of [illegible], de-

[illegible] et Ranis. [illegible]bibat, &c. [illegible]
[illegible]guarus, [illegible]iri, sed non [illegible]
[illegible]rausâ freq[illegible]

Digitized by Google

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt **Date of investigation**: 6/17/08

Book title: The Anatomy of Melancholy, volume III
Author: Robert Burton
Year of Publication: 1875 by W. J. Widdleton Publishers. Original was published in 1661.
Place of Publication: New York
ISBN number?:
Content/Disciplinary area: medical history, first attempt to describe depression and other ailments.

AVAILABILITY SEARCH

Source of scanned book:	Google Books.
Comments on condition/provenance of book:	This is the second and only other version in Google Books that has full view available. 11 editions are on Google Books, but not all with viewing capability.
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	This is book is mistitled by Google Books- it comes up as "The Anatomy of melancholy v. 1" but is actually volume III. One other text in Google Books, with the same time, actually is the first volume.
MARC record available?	No
Can the book be searched through an OPAC? If so, which one?	Yes – GEORGE's catalog has different online versions, all made available through EEBO (Early English Books Online). These are digitized copies of the 1661 original and other editions- all in one file (no volumes).
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	This book has a wonderful index in the back, with blue hyperlinks that will direct the researcher right to the page of the indexed item. (For example, if one wants to reference a 'community of wives a cure of jealousy' one may look in the c section of the index, and click on the link after this item – it will take you directly to page 331 in the book. This feature is fantastic. (See attached picture)

SCAN SEARCH

search paradigm settings (example: every 10 double-pages) every 10 pages, scrolled through the pages so as to not miss any details.

Page(s):	readability/ quality of scan	anomalies?	other?	other?
1, 10, 12, 14, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 44, 46, 48, 52, 54, 56, 58, 60, 62, 66, 68, 70, 72, 76, 78, 80, 81, 82, 84, 85, 87, 88, 89, 91, 92, 93, 95, 97, 98, 100, = 200, 201 (I	excellent excellent text …	none these pages have three of the scanner's fingers (sometimes one or two fingers) on the bottom left corner—Google has imposed white squares on them to cover up their shape. Few do not have white squares imposed and reveal all of the fingers with detail.	(see attached picture as an example)	(Pages 30, 34 is particularly tacky with the finger, finger nail, etc.)

could go on)				
70	good	the top of page 70 was moved while scanning, text is stretched.	This does not appear to affect OCR search of words/letters.	
76	poor	This page is almost at a diagonal – was completely moved during the scan	Makes the majority of the text on this page illegible or not searchable.	This page is an anomaly. (see attached picture of Page 76).
77	Page 77 has part of page 76 covering it, making some of the text in upper right illegible			
104	Page 104 has a great deal of blurred portion of page 105 in it's scan	This does not impede text however		
210	excellent	none	fingers have been whited out	
300	excellent	none		
350, 500	excellent	none	fingers whited out	
524	excellent	This is not text- Google scanned the back library card section of the book		

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	Comment (Is OCR search permitted): Y / N
purpurissa	This word was in an end note- very small. OCR search did not detect it.
suas	This word was in the same end note as above word (page 79), and was detected.
honestiores	This word was in the same end note as above two words, and was not detected.
apogæo	I saw this word a few times- search did not detect it. Searches for 'apogaeo' and 'apog*' were also unsuccessful. Symbols are problematic.
love	highlighted all finds in yellow - successful search.
exercise	highlighted all finds in yellow - successful search.
ache	highlighted all finds in yellow - successful search.
stomach	highlighted all finds in yellow - successful search.
birth	highlighted all finds in yellow - successful search.
incarnatus	highlighted all finds in yellow - successful search.

GENERAL OBSERVATIONS

Comment on ease of discovery in integrated interface (OPAC or better)	Very
Overall quality of scan of this book	The text scan quality is excellent – except for some pages that were moved during scan. However, the OMNIPRESENCE of a scanner's FINGERS in this book is really daunting- it is distracting, takes away from the book (it's unpleasant to look at)
Ease of access to full text	Yes
Ability to copy, quote, cite?	Yes- possibly from a downloaded version
Ability to search?	Yes
Ability to download?	Yes!
Ability to print?	Yes

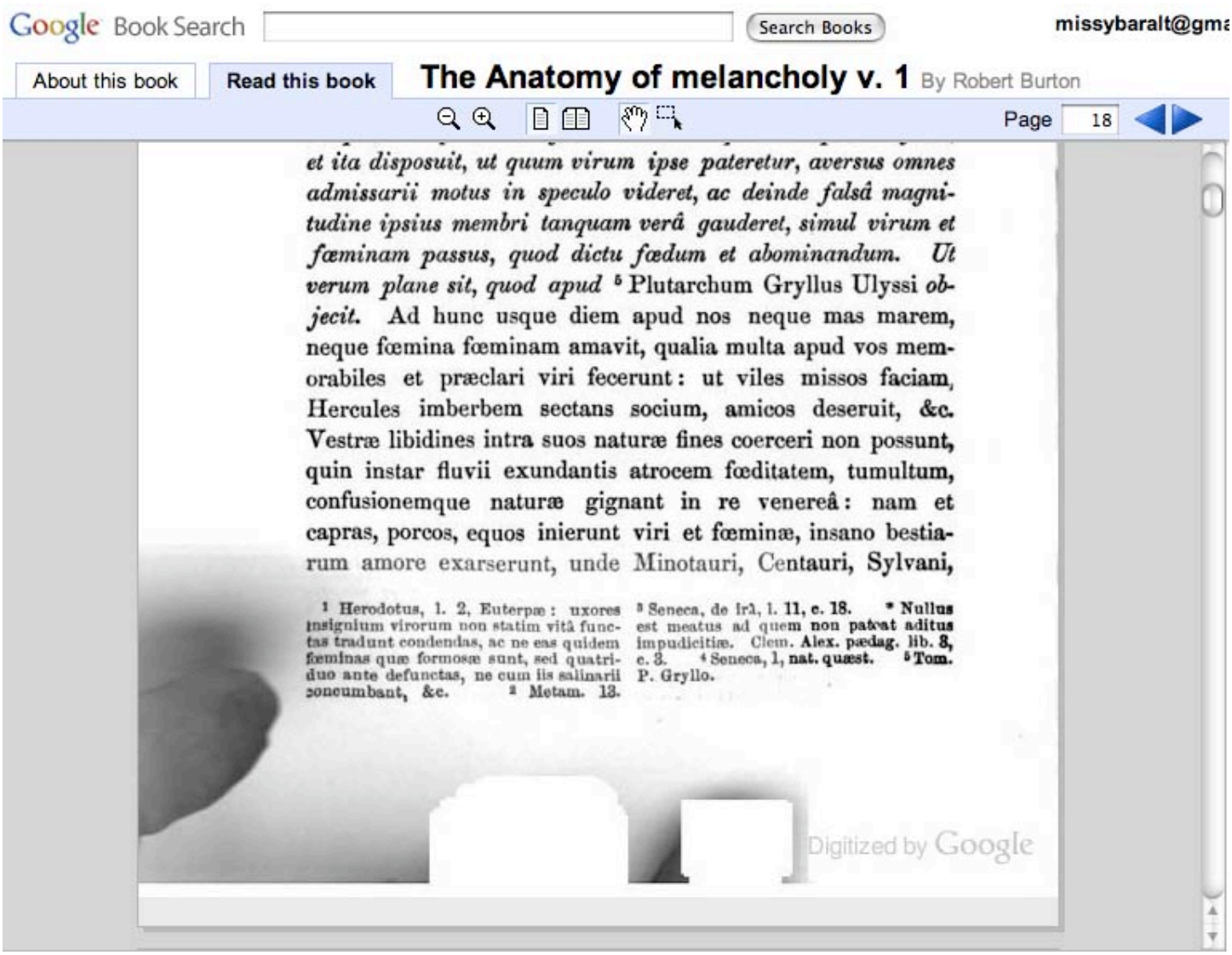
CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	Good- minus some Latin characters and small text that exists in footnotes. Also, pages that were moved during the scan have text that is not readable by OCR or the researcher.
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	It could be, if not for so many instances of a scanner's fingers on the pages.
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Must get rid of fingers.
Issues with this book that would affect usefulness for scholars	OCR does not find everything- this is certainly an issue.
Can this book be shared with other scholars/students?	Yes, by emailing the link.
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	Volume 2 is apparently missing from Google's online Book collection. This is Volume 3, but is mislabeled as a Volume 1 in Google Book.
Other notes	The number of pages with fingers visible is somewhat appalling.

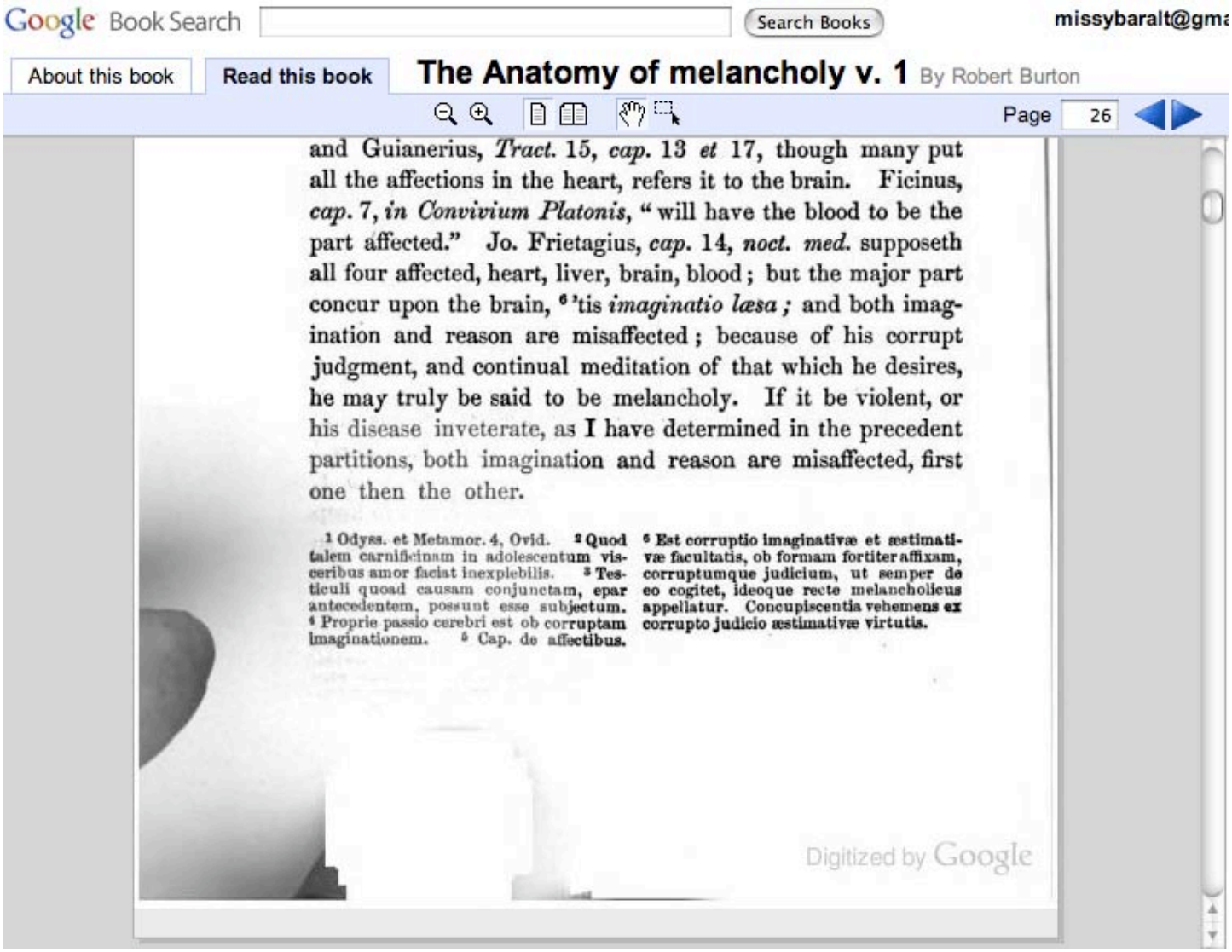
Book URL:
http://books.google.com/books?id=bEUJAAAAIAAJ&printsec=frontcover&dq=editions:0YP4Pr8J8AP6uDLt5yNC#PPA210,M1

Date accessed: 6/17/08 – 6/18/08

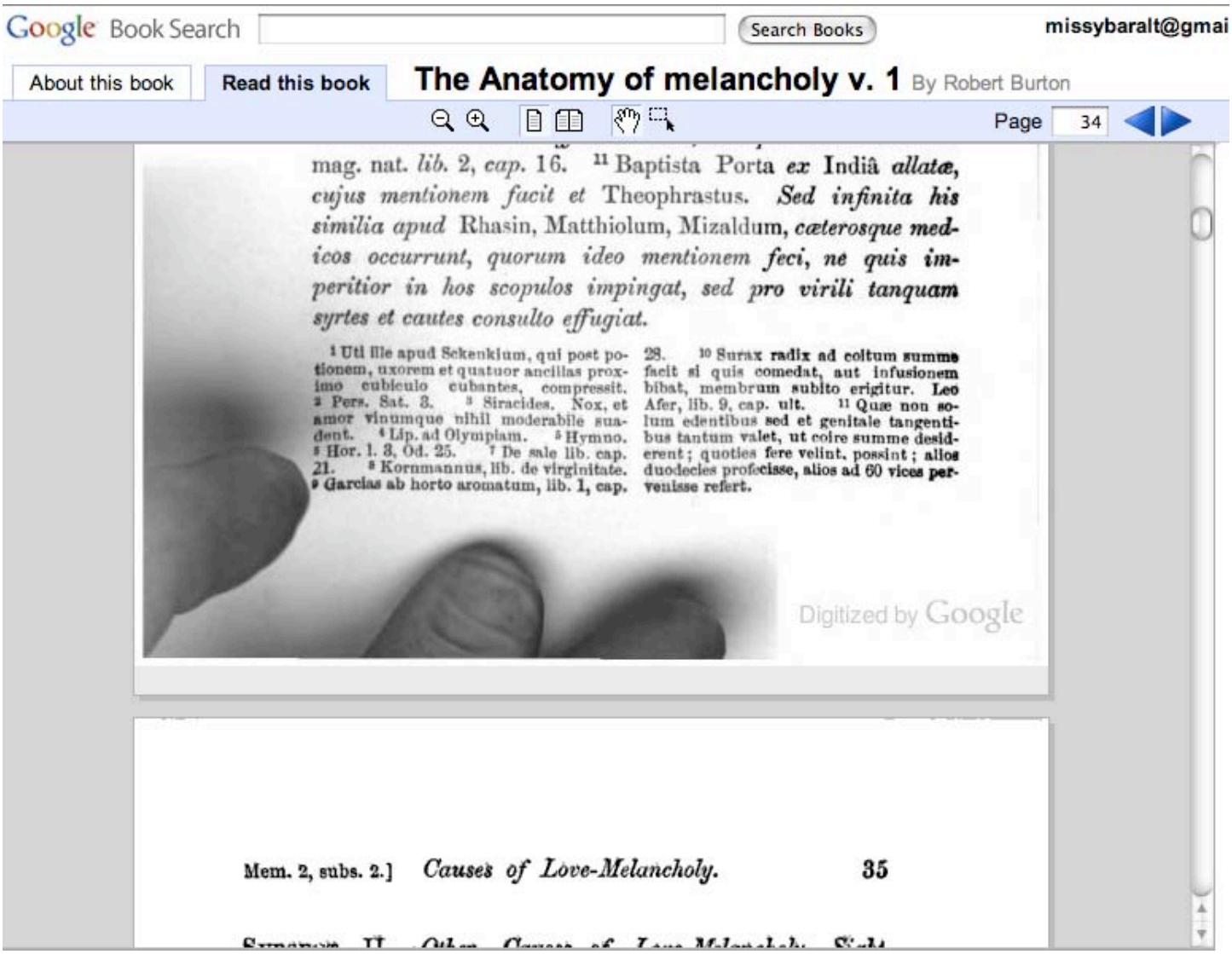
Example of scanner's fingers in scan, with while boxes imposed over the fingers by Google



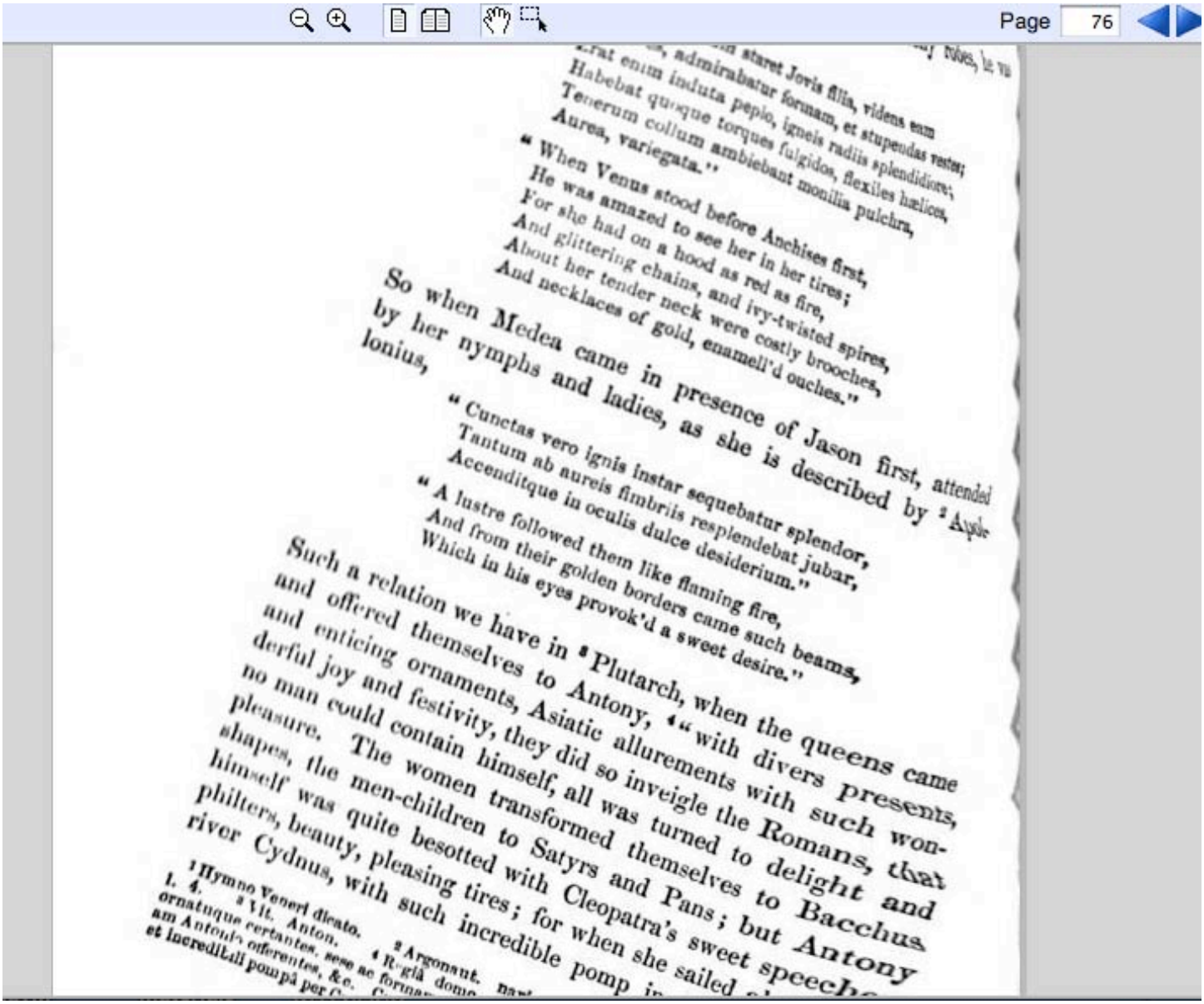
Second example of whited-out fingers



Not attempt to white-out these fingers (page 34)



Page 76



...staret Jovis filia, videns eam
...admirabatur formam, et stupendas vestes;
Erat enim induta pepio, igneis radiis splendidiore;
Habebat quoque torques fulgidos, flexiles hælices,
Tenerum collum ambiebant monilia pulchra,
Aurea, variegata."

"When Venus stood before Anchises first,
He was amazed to see her in her tires;
For she had on a hood as red as fire,
And glittering chains, and ivy-twisted spires,
About her tender neck were costly brooches,
And necklaces of gold, enamell'd ouches."

So when Medea came in presence of Jason first, attended by her nymphs and ladies, as she is described by [2] Apollonius,

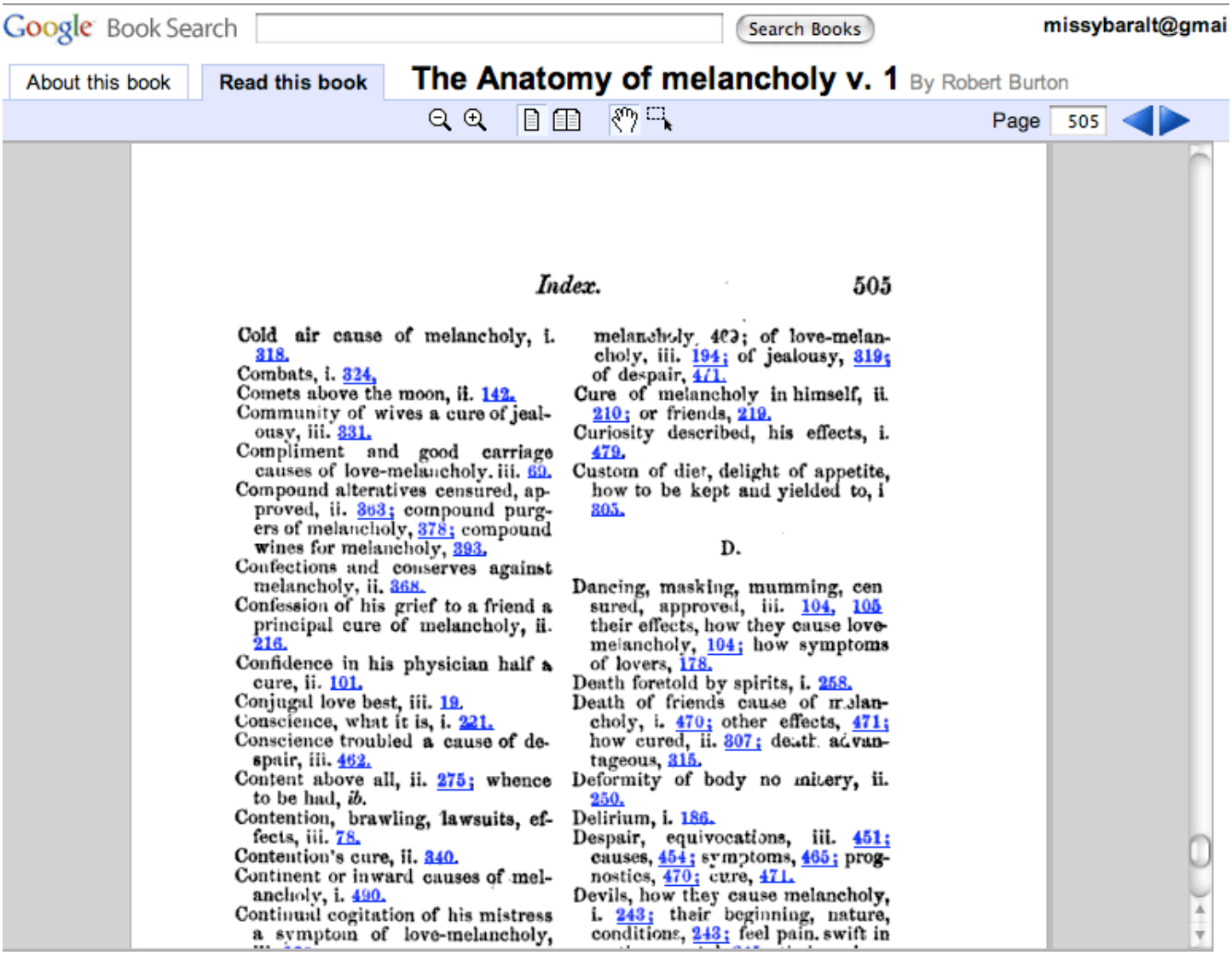
"Cunctas vero ignis instar sequebatur splendor,
Tantum ab aureis fimbriis resplendebat jubar,
Accenditque in oculis dulce desiderium."

"A lustre followed them like flaming fire,
And from their golden borders came such beams,
Which in his eyes provok'd a sweet desire."

Such a relation we have in [3] Plutarch, when the queens came and offered themselves to Antony, [4] "with divers presents, and enticing ornaments, Asiatic allurements with such wonderful joy and festivity, they did so inveigle the Romans, that no man could contain himself, all was turned to delight and pleasure. The women transformed themselves to Bacchus shapes, the men-children to Satyrs and Pans; but Antony himself was quite besotted with Cleopatra's sweet speech... philters, beauty, pleasing tires; for when she sailed ... river Cydnus, with such incredible pomp in...

[1] Hymno Veneri dicato. [2] Argonaut. ... l. 4. [3] Vit. Anton. [4] Regiâ domo ... ornatuque certantes, sese ac forma... am Antonio offerentes, &c. ... et incredibili pompâ per C...

Example picture of index with hyperlinks



Index. 505

Cold air cause of melancholy, i. 318.
Combats, i. 324.
Comets above the moon, ii. 142.
Community of wives a cure of jealousy, iii. 331.
Compliment and good carriage causes of love-melancholy. iii. 69.
Compound alteratives censured, approved, ii. 363; compound purgers of melancholy, 378; compound wines for melancholy, 393.
Confections and conserves against melancholy, ii. 368.
Confession of his grief to a friend a principal cure of melancholy, ii. 216.
Confidence in his physician half a cure, ii. 101.
Conjugal love best, iii. 19.
Conscience, what it is, i. 221.
Conscience troubled a cause of despair, iii. 462.
Content above all, ii. 275; whence to be had, *ib.*
Contention, brawling, lawsuits, effects, iii. 78.
Contention's cure, ii. 340.
Continent or inward causes of melancholy, i. 490.
Continual cogitation of his mistress a symptom of love-melancholy, melancholy, 469; of love-melancholy, iii. 194; of jealousy, 319; of despair, 471.
Cure of melancholy in himself, ii. 210; or friends, 219.
Curiosity described, his effects, i. 479.
Custom of diet, delight of appetite, how to be kept and yielded to, i 805.

D.

Dancing, masking, mumming, censured, approved, iii. 104, 105 their effects, how they cause love-melancholy, 104; how symptoms of lovers, 178.
Death foretold by spirits, i. 258.
Death of friends cause of melancholy, i. 470; other effects, 471; how cured, ii. 307; death advantageous, 315.
Deformity of body no misery, ii. 250.
Delirium, i. 186.
Despair, equivocations, iii. 451; causes, 454; symptoms, 465; prognostics, 470; cure, 471.
Devils, how they cause melancholy, i. 243; their beginning, nature, conditions, 243; feel pain, swift in

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Baralt Date of investigation: 6/15/08

Book title: Syntactic Structures
Author: Noam Chomsky
Year of Publication: **This book has four different editions:** The Hague: Mouton, 1957. Reprint. Berlin and New York, 1985; Berlin and New York: Mouton de Gruyter, 2002. Translated as Structures Syntaxiques. France: Editions du Seuil, 1994. Also appears to have several reprinted versions, for example, Walter de Gruyter of Berlin and New York is the copyrighted version on Google Books with limited public access.
Note: our library at Georgetown has no electronic version. Blackwell's Book Services does allow an 'inside-the-book' view of a table of contents only.
Place of Publication: see above.
ISBN number: 3-11-017279-8
Content/Disciplinary area: linguistics
browser and computer used for this search: Safari, Mac

AVAILABILITY SEARCH

Source of scanned book:	Google Books Google Books version: Walter de Gruyter Inc. 1978
Comments on condition/provenance of book:	color copy scan of front cover. Given that it's copyrighted, several pages missing: 10, 16-17, 23-24, 30-31, 37-38, every page after page 42!
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	straight link to book (de Gruyter's) via Google Book search. R.E. metadata: Google does provide an 'other editions' link – which leads to 16 other versions/editions. Several seem to be repeats- for example, a U of Michigan version, digitized Feb. 22, 2008, is listed four times. None of these 'other editions' are available for view
MARC record available?	No (cannot find this information) any ideas on where to find this?
Can the book be searched through an OPAC? If so, which one?	No – at least not available through Georgetown library or WorldCat
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	Table of contents is provided, but because of limited page access and search ability of book, one cannot go to certain pages. Navigating is not possible- must scroll through book (or those pages which are available) to get to desired page. Must scroll back up to return to table of contents. Tedious.

SCAN SEARCH

search paradigm settings (example: every 10 double-pages) Due to limited number of pages available, I looked at each page provided online in this book

Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
front cover	excellent – in color	none		
table of contnts	good	none	no page number?	
11-15, 18	good	none		
19	good	none	graphics clear	
21-22	good	none	footnote clarity (ex. pg. 22) can be improved by zooming in on document	
25-29	good	none		
32	good		Subscript characters slightly difficult to read – (19) (v). By contexts one can deduce subscripted text unv means 'unvoiced consonant', but scan is blurred.	
33-36	good	none		

39	good		again, subscripted text taxing to read
40-42	good	none	

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	Comment
	(Is OCR search permitted): Y / N I think so – let's ask JO'D about this
grammar	search results tell me those pages on which 'grammar' appears, but are not viewable. Says "Page viii You have reached your viewing limit for this book" or "Page 93 You have reached your viewing limit for this book." Frustrating. Does not tell me number of hits, but gives me all results. Highlights word in yellow throughout text. Footnote on page 36 divides gramm-ar into two lines, is the only case I could find where yellow highlighting didn't cover whole word. Otherwise, search results well done.
constituent	takes a few clicks to get yellow highlighting of search results to change from 'grammar' to 'constituent.' In fact, some pages still have 'grammar' highlighted in yellow, some don't. Seems I must refresh page to clear results and begin new word search. Good- I like that Google gives me results with a full sentence that includes the word. On page 36, search yielded constituent in footnote – good. Again, I cannot clear yellow highlighting of my previous search, 'grammar' on some pages.
paradigm	Only one hit, on a page not accessible to me "Page vi You have reached your viewing limit for this book." Now yellow highlighting of both 'grammar' and 'constituent' is present.
unv	I was curious if the blurriness of this subscripted text, unv [unvoiced consonant] I found on page 32 would be readable in search. **In fact, OCR search did NOT find the subscripted acronym. No results, even though it is present three times on page 32.**
sing	This was a subscripted text example for Noun Phrase used by Chomsky often (see page 29). Ex: NP_{sing} **OCR search did NOT find it.**
pl	Also a subscripted text example (see page 29). **OCR search did NOT find it.**
morphophonetics	2 hits – one on page 32 in italics, successful find. The other is on page not accessible.
Ø	Symbol for null – used often in linguistics [for example, to indicate null subject position for those languages that don't require this syntactic position be overtly filled]. **OCR search did NOT find it,** even though it appears at least four times in available text. (Examples: twice in footnote on page 29, twice in regular text on page 39). [As a linguist I'm shocked I can't search for this!]
∑	Interesting- the sigma symbol, also important in linguistics – yield 30 search results, but OCR reader did not highlight anything in yellow. On most of the pages with ∑, another symbol is present, §. Possible confusion with the two? (Though neither is highlighted). In fact, sigma appears 3 times on page 29, page 32, twice on page 40, twice on page 41, once in footnote, page 41.
S	Search results successful, highlighted in yellow. (S = sentence in linguistic description; example, "Let S_1, S_2, S_3 … be declarative sentences in English", etc.). 22 hits total, all successful.

GENERAL OBSERVATIONS

Comment on ease of discovery in integrated interface (OPAC or better)	Difficult- not available in Georgetown or WorldCat. Google search results indicate U Michigan has several digitzed versions- these unavailable to me as common layman. Only version I could access was limited-access version in Google Books.
Overall quality of scan of this book	Good
Ease of access to full text	No – limited page access, only goes to page 42.
Ability to copy, quote, cite?	No – I cannot highlight and copy text
Ability to search?	OCR word search, yes. However linguistic and IPA (International Phonetic Alphabet) symbols are problematic.

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Ability to download?	No.
Ability to print?	No
Scanned book allow for consulting of contents?	No
How successful is OCR recognition?	For words, Roman alphabet characters – successful. For IPA symbols, not so.
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Yes.
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	I'd rate it at 85%.
Issues with this book that would affect usefulness for scholars	In digitized format, my linguistic searches for subscripted text and IPA symbols is not possible. This is a major issue for linguists. Also, scrolling capability is frustrating- much like a papyrus role. I cannot jump from, say page 12 to 40, for example. I must scroll through entire document.
Can this book be shared with other scholars/students?	Yes, by emailing the URL in a link. However, it is a copyrighted, Google limited access book.
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	I would say no.
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	No full-text access to any of the other editions on the web.
Other notes	Very surprised to see that searches for subscripted text and phonetic symbols do not yield any results in a book on linguistics by Noam Chomsky, given their ubiquitous nature in such a text.

URL for this book:
http://books.google.com/books?id=a6a_b-CXYAkC&dq=syntactic+structures&pg=PP1&ots=V7KrC3egEr&sig=hoVEMtNqv5AKCpa4TJhbN4on2Vw&hl=en&prev=http://www.google.com/search%3Fhl%3Den%26q%3Dsyntactic%2Bstructures%26btnG%3DGoogle%2BSearch&sa=X&oi=print&ct=title&cad=one-book-with-thumbnail#PPP1,M1

accessed 6/15/08

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt	**Date of investigation**: August 28

Book title: Crónicas Potosinas costumbres de la edad medieval Hispano-Americana
Author: Vicente Quesada
Year of Publication: 1890
Place of Publication: Paris: Biblioteca de la Europa y América
ISBN number?: NA
Content/Disciplinary area: Potosí, literature, Hispano-America

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	Original provided by New York Public Library
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Easily accessible in Google Books. However, this is a second version, and no information is given on first version
MARC record available?	NA
Can the book be searched through an OPAC? If so, which one?	No
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	Metadata provided sufficient.

SCAN SEARCH

search paradigm settings (example: every 10 double-pages)	scrolled through entire book

Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
No major anomalies in this book!				

WORD SEARCHES DONE WITH THIS BOOK
(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	**Comment**
	(Is OCR search permitted): Y / N
indios	successful detection

amabais	successful detection
Trabóse	successful detection
"lluvia de sangre"	successful detection
satisfacción	successful detection
Caricarí	successful detection
carruajes	successful detection
fuese	successful detection
levantóse	successful detection
adivináis	successful detection

IMAGE ASSESSMENT	
If the original text includes graphics/illustrations, are they included in the electronic version?	No images present in this text
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	
Is the scan on a gray-scale or color scale?	
Quality – does it look like the real thing?	
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	
Other notes	

GENERAL OBSERVATIONS	
Comment on ease of discovery in	Very easily accessible in Google Books

integrated interface (OPAC or better)	
Overall quality of scan of this book	Good
Ease of access to full text	Yes
Ability to copy, quote, cite?	Yes
Ability to search?	Yes
Ability to download?	Yes
Ability to print?	Yes
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	seemingly perfect – no issues! All successful detections
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Yes, absolutely
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes
Issues with this book that would affect usefulness for scholars	NONE
Can this book be shared with other scholars/students?	Yes, by emailing the URL
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	Only one comment - this is the second version, but no information is given on first version
Other notes	

Book URL:
http://books.google.com/books?id=rhETAAAAYAAJ&pg=PA329&dq=arsans&lr=lang_es&as_brr=1&vq=%22hasta+este+punto+que+es+el+último+de+mi+vida,+en+el+cual+me+arrepiento+de+lo+hecho,+y+pido%22&source=gbs_quotes#PPT1,M1
Date accessed: August 28, 2008

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt	**Date of investigation**: August 21, 2008

Book title: Die Alte Landschaft Babylonien nach den Arabischen Georgraphen
Author: Dr. Maximilian Streck
Year of Publication: 1900
Place of Publication: Buchhandlung und druckerei E. J. Brill
ISBN number?: NA
Content/Disciplinary area: Iraq history and geography from 1900

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	Original text provided by the University of Michigan
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Easily accessible in Google Books. Metadata provided is limited.
MARC record available?	No
Can the book be searched through an OPAC? If so, which one?	No
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	All of table of contents is hyperlinked in the book

SCAN SEARCH

search paradigm settings (example: every 10 double-pages)	scanned through entire book

Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
132, 134, 136, xi (where pagination starts over??), 176, 190,	top corner is whited out, eliminating some of the text			
135	page was moved during scan, making reading/ searching impossible	see figure 1		
139	text is so blurred on this page it is rendered illegible			
x (Google recognized erroneously as xi by Google pagination), xi	large white strip through middle of page eliminating text			
199	top half of page completely whited out			

286				

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	Comment
	(Is OCR search permitted): Y / N
□□	unsuccessful detection (present on page 82) **df**
Brücke	successful detection
□□□□□	unsuccessful detection (present on page 82) **df**
quatiᶜa	unsuccessful detection (on page 108)
askar el-Mahdi	unsuccessful detection (on page 111)
Ansiedelungen	successful
mûs ā	unsuccessful detection (on page 116)
Mu ʿ izz	unsuccessful detection (on page 153)
Stadt	successful
Qorans	successful

IMAGE ASSESSMENT

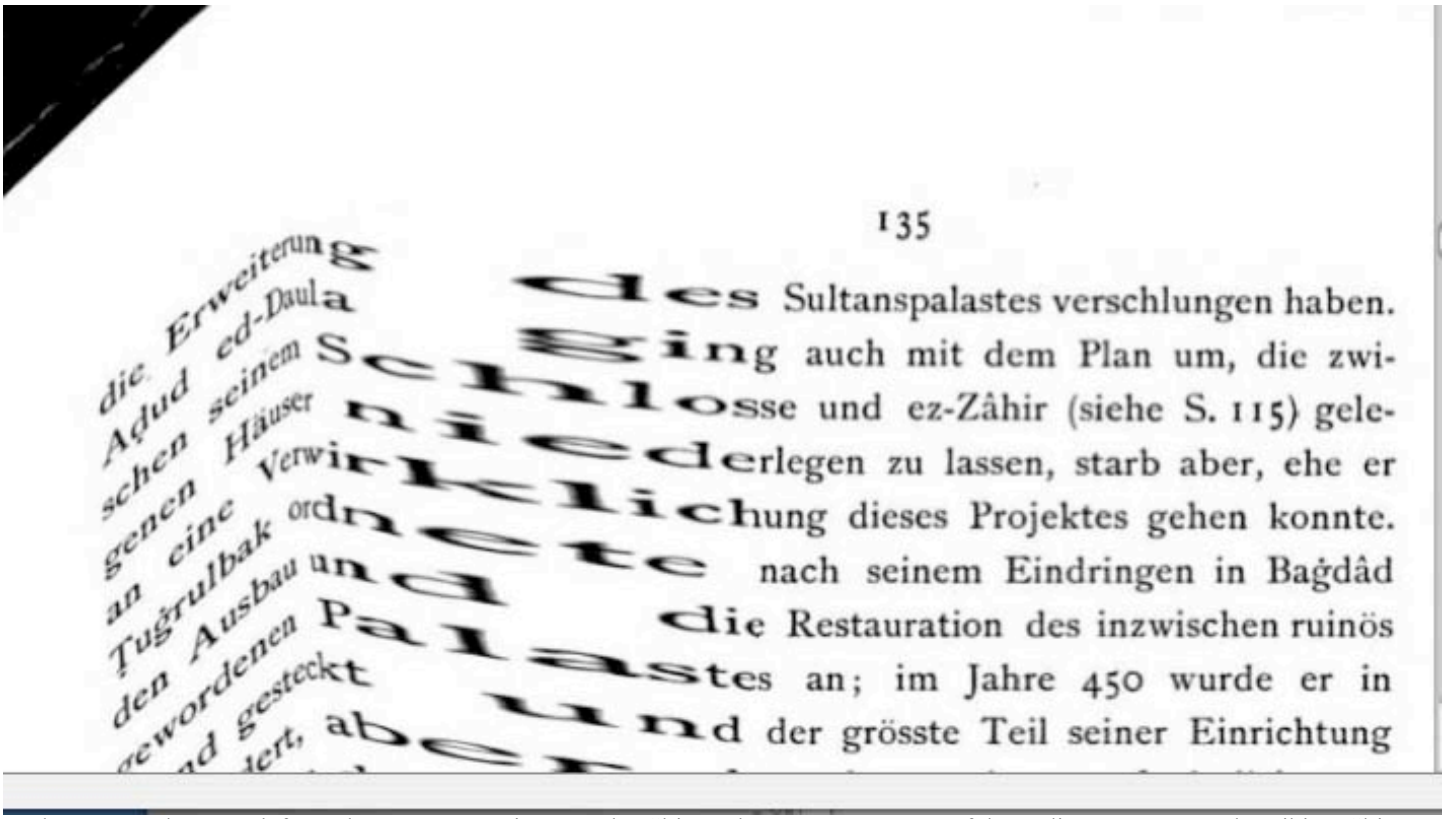
If the original text includes graphics/illustrations, are they included in the electronic version?	No images in this book
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	
Is the scan on a gray-scale or color scale?	
Quality – does it look like the real thing?	
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming	

capabilities, filter options)	
Other notes	

GENERAL OBSERVATIONS	
Comment on ease of discovery in integrated interface (OPAC or better)	Easily accessible in Google books
Overall quality of scan of this book	Poor – too many pages were moved during scan and/or have whited out sections
Ease of access to full text	Yes, besides whited out sections that preclude reading of text
Ability to copy, quote, cite?	Yes
Ability to search?	Yes
Ability to download?	Yes
Ability to print?	Yes
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	Poor- no reading of Arabic, Hebrew, or other alphabets present in this book
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Only if re-scanning of pages with anomalies
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes
Issues with this book that would affect usefulness for scholars	No reading/ searching permitted of words in other alphabets or even of some letters with distinct accent marks not found in English
Can this book be shared with other scholars/students?	Yes, by emailing the URL
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	NA
Other notes	Right before page 173 there is a restart of the pagination system by Google – many of these pages are illegible

Book URL: http://books.google.com/books?id=pr0LAAAAIAAJ&pg=RA1-PA286&dq=History+of+Iraq+date:1700-1902&lr=lang_ar&as_brr=0#PPA135,M1
Date accessed: August 21, 2008

Figure 1 page 135 example of page being moved during scan



Today I wanted to search for early grammars written on the Chinese language – or some of the earliest attempts at describing Chinese grammar pedagogy. According to Xing (2006), some of the earliest books on this topic were written by mostly Western missionaries - many used Latin to describe Chinese. The earliest books include Matteo Ricci, 1580; João Rodriguez, 1604; Francisco Varo, 1703; Joseph de Prémare, 1831; Giles, 1872; Hiller, 1886; Gibson, 1887 (Xing pg. 63). I attempted to look for all of these volumes in Google Books.

Matteo Ricci, 1580 - search yielded no results.
João Rodriguez, 1604 – search yielded no results.
Francisco Varo, 1703 – Google Books lists it (arte de la lengua mandarina), but it is not available for view.
Joseph de Prémare, 1831 – Google Books lists it, but it is not available. (Notitia linguae sinicae). 2 other versions are listed but are also not available. There is one full-view version, but it is an edition that was translated into English. Not too much metadata is provided for this book. (Georgetown has this book in off-campus storage).

However, another title by this author, Vestiges des principaux dogmes chrétiens, tirés des anciens livres chinois, is available for full view. I shall do this for my search today – I am curious to see how a book before 1900 with Chinese text is recognized.

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt **Date of investigation**: 6/28/08 - 6/29/08

Book title: Vestiges des principaux dogmes chrétiens, tirés des anciens livres chinoi
Author: Joseph de Prémare
Year of Publication: 1878
Place of Publication: Bureau des Annales de philosophie chrétienne
ISBN number?: NA
Content/Disciplinary area: Chinese language, early grammar

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	Digitized by Google on Sept. 13, 2007. Original is from Harvard University.
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Not too much is provided – I had to do an author search in Google Books.
MARC record available?	NA
Can the book be searched through an OPAC? If so, which one?	No. In fact, Georgetown does not have this book – WorldCat indicates that only 14 books worldwide own it. Wonderful to have a digitized version provided online with Google.
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	Pages 502, 505-508 (table of 'articles') have wonderful hyperlinked items that the researcher may click on, and be sent directly to article (or chapter point).

SCAN SEARCH

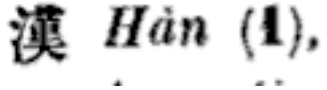
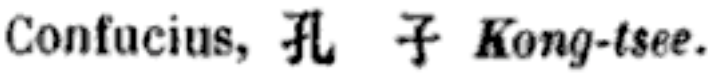
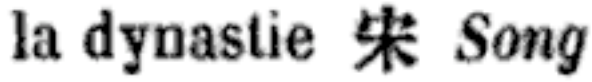
search paradigm settings (example: every 10 double-pages) Scanned through entire book

Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
inner title page	entire left side of detailed border, done with Chinese characters, is left out of the scan. This may be because spine was not opened enough.			
11	small text on this page, some words not detected by OCR (for example une àutre)			
14	text on this page close to spine is slightly stretched from movement during scan			
248	THIS PAGE WAS MOVED DURING SCAN – page lost/illegible.	See Image II.	(In fact, it seems that a page turn from 246-247 was being done- page 248 was completely left	

			out).	
249	THIS PAGE WAS MOVED DURING SCAN – page lost/illegible.	See Image III.		
280	text at top of this page just slightly blurry			
402	This page is actually present three times- the first is correct, the second and third pages '402' are simply blank white pages. Odd.	See Image IV.	Google could just eliminate 402(b), 402(c), as pagination begins again perfectly on page 403.	Not sure if in original there were two blank folios/pages between 402 and 403?
404	This page was moved slightly during scan – text is in a wave shape at top of page.	It appears this does affect OCR recognition of words that are misshaped.	See Image V	
432, 433	These pages are MISSING.	Google pagination names page 431 (actually present) as 433. Google pagination goes from 430 to 433.	Pagination realigns correctly on page 434.	
	Besides the above anomalies, the remaining pages in this book, up to page 519, are in excellent, almost flawless condition.			

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	Comment (Is OCR search permitted): Y / N
Chine	Some successful, some not. For example, *Chine* is in a footnote on page xiv, and was not successfully detected by OCR.
chinoise	Search /recognition successful.
漢	This symbol means Han (as in Han Dynasty). Search result Unsuccessful (it is on page 23 in book). 
hiéroglyphiques **孔子**	Unsuccessful in footnote on page 27, but successful recognition on pages 43, 50 and 201 – all presented to the researcher in yellow highlighting. this symbol means Confuscius (note- this book uses traditional as opposed to simplified Chinese symbols. OCR recognition UNSUCCESSFUL. (Note- interesting to see French phonetic spelling of the word, "Kong-tsee" as compared to Pinyin "kông zi")  see book page 27
l'empereur **宋**	successful recognition This symbol means Song (as in Song Dynasty). It is on page 42. Recognition NOT successful. 
Se-tà-ky-chou	This word was in a footnote on page 45. It was detected, however, note how the OCR recognized the Chinese symbols immediately afterwards:

"Il a laissé des commentaires et un ouvrage intitulé : **Se-tà-ky-chou** " |7tj *-fc •fif ^9 'ca * merveilleuses productions. **...**"

Here is the real text from the book:

Se-tà-ky-chou 四 大

內 symbol for internal – present on page 74. Recognition NOT successful.

內 *nuei*, interne,

陰 symbol for yin. Search NOT successful.

yn 会

赤 symbol for red – present in a large chart at an angle on page 132. OCR Recognition NOT successful.

大 symbol for big, present on page 161. Search result NOT successful.

explication Word present in a graph on page 177, going in an upwards direction. Search results NOT successful.

月 Symbol for moon (la lune), present on page 218. Search results NOT successful.

et 月 *yue* la lune.

土 symbol for land, present on page 258. Search results NOT successful.

士 *la terre.*

動 symbol for movement, present on page 309 and 310. Search results NOT successful.

動
le mouvement,

女 symbol for woman, present on page 437. Search results NOT successful.

Niu 女 *femme,*

I needed to do searches for as many Chinese characters as possible (and for those which I could look up via French translations). I am now completely convinced that there is no OCR recognition for Chinese characters in this book.

IMAGE ASSESSMENT

If the original text includes graphics/illustrations, are they included in the electronic version?

There is a small image on page 283 – I am not sure what it is.

If there are no graphics/illustrations are

No

there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	NA
Is the scan on a gray-scale or color scale?	NA
Quality – does it look like the real thing?	NA
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	NA
Are the illustrations searchable?	No
Ability to print illustrations? (is there a watermark?)	Yes, if the researcher downloads entire document.
Are the images downloadable? (file size, pixel quality)	Within the text as a whole, yes.
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	No
Other notes	There are also some hieroglyphs present in this book, for example on page 305, and figures on page 376 (consisting mostly of lines and circles).

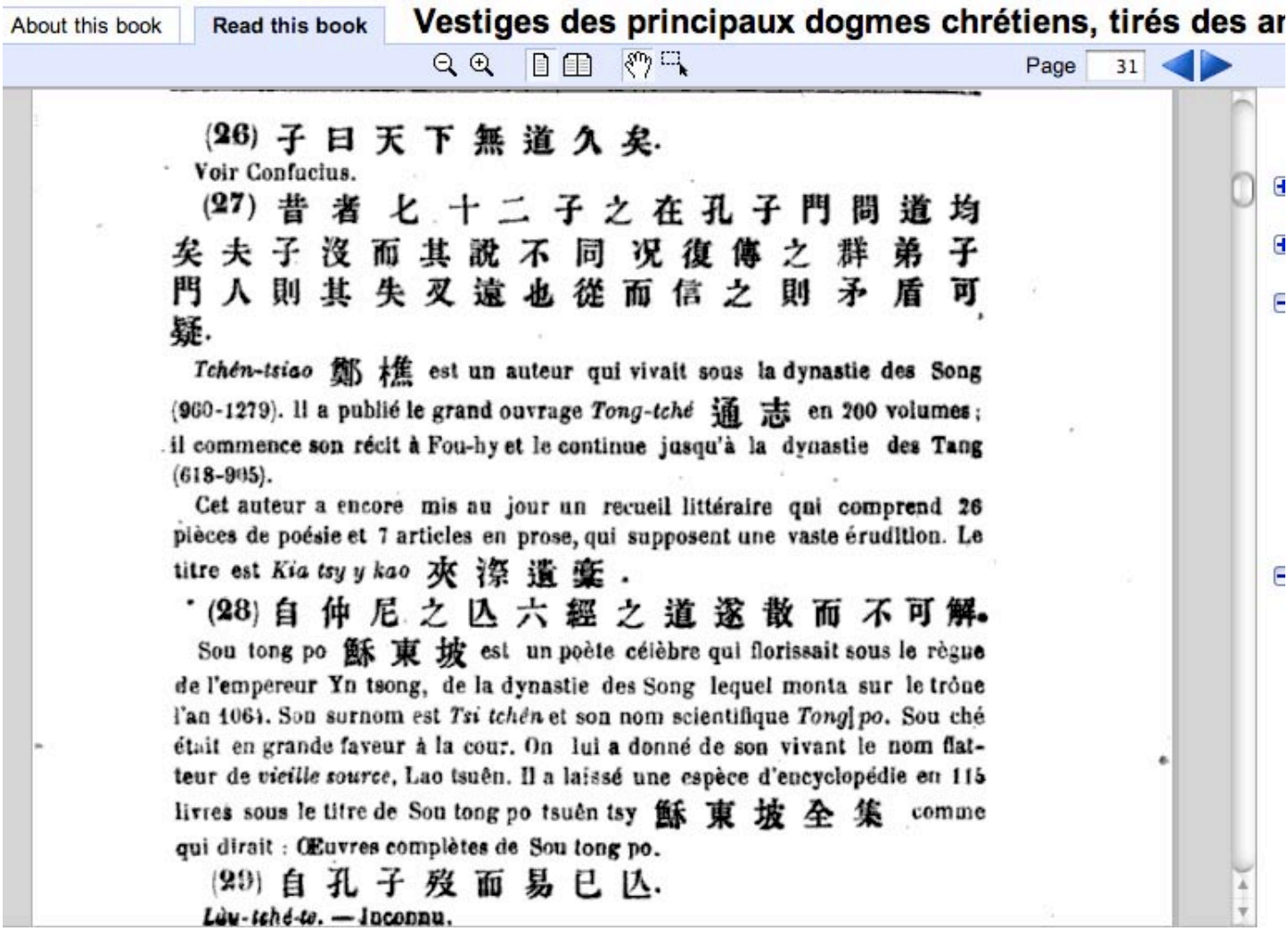
GENERAL OBSERVATIONS

Comment on ease of discovery in integrated interface (OPAC or better)	Very, though Google Books had the only digital copy. Not many libraries worldwide own this text.
Overall quality of scan of this book	Excellent, besides some missing pages and scan anomalies.
Ease of access to full text	Very
Ability to copy, quote, cite?	No
Ability to search?	Yes – except for Chinese characters. Given that Chinese characters is the subject matter of this book, this is a significant downfall to the present digitization project.
Ability to download?	Yes
Ability to print?	Yes
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	For French words- good, except some small text in footnotes. For Chinese characters – impossible.
Worthy as archive? (i.e., would this copy be acceptable as an archival copy	Definitely.

of the book if physical copies were lost or destroyed?	
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Absolutely – fantastic quality overall.
Issues with this book that would affect usefulness for scholars	No Chinese character searching. The hyperlinks provided in the back of the book are wonderful.
Can this book be shared with other scholars/students?	Yes, if link is emailed.
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	NA
Other notes	Def necessary to research Pinyin and Chinese character writing (in digital format) for this book.

Book URL:
http://books.google.com/books?id=E6EGAAAAQAAJ&pg=PA8&dq=inauthor:Joseph+inauthor:de+inauthor:Prémare&lr=&as_brr=0#PPP1,M1
Date accessed: 6/29/08

The Chinese characters in this book are really beautiful, This is an example page:



(26) 子曰天下無道久矣.

Voir Confucius.

(27) 昔者七十二子之在孔子門問道均矣夫子沒而其說不同況復傳之群弟子門人則其失又遠也從而信之則矛盾可疑.

Tchén-tsiao 鄭樵 est un auteur qui vivait sous la dynastie des Song (960-1279). Il a publié le grand ouvrage *Tong-tché* 通志 en 200 volumes; il commence son récit à Fou-hy et le continue jusqu'à la dynastie des Tang (618-905).

Cet auteur a encore mis au jour un recueil littéraire qui comprend 26 pièces de poésie et 7 articles en prose, qui supposent une vaste érudition. Le titre est *Kia tsy y kao* 夾漈遺稾.

(28) 自仲尼之亾六經之道遂散而不可解.

Sou tong po 蘇東坡 est un poète célèbre qui florissait sous le règne de l'empereur Yn tsong, de la dynastie des Song lequel monta sur le trône l'an 1064. Son surnom est *Tsi tchén* et son nom scientifique *Tong] po*. Sou ché était en grande faveur à la cour. On lui a donné de son vivant le nom flatteur de *vieille source*, Lao tsuên. Il a laissé une espèce d'encyclopédie en 115 livres sous le titre de Sou tong po tsuên tsy 蘇東坡全集 comme qui dirait : Œuvres complètes de Sou tong po.

(29) 自孔子歿而易已亾.

Lèu-ché-to. — Inconnu.

Image II

Page 248 – lost unfortunately, due to a scan anomaly. Page was being turned during scan.

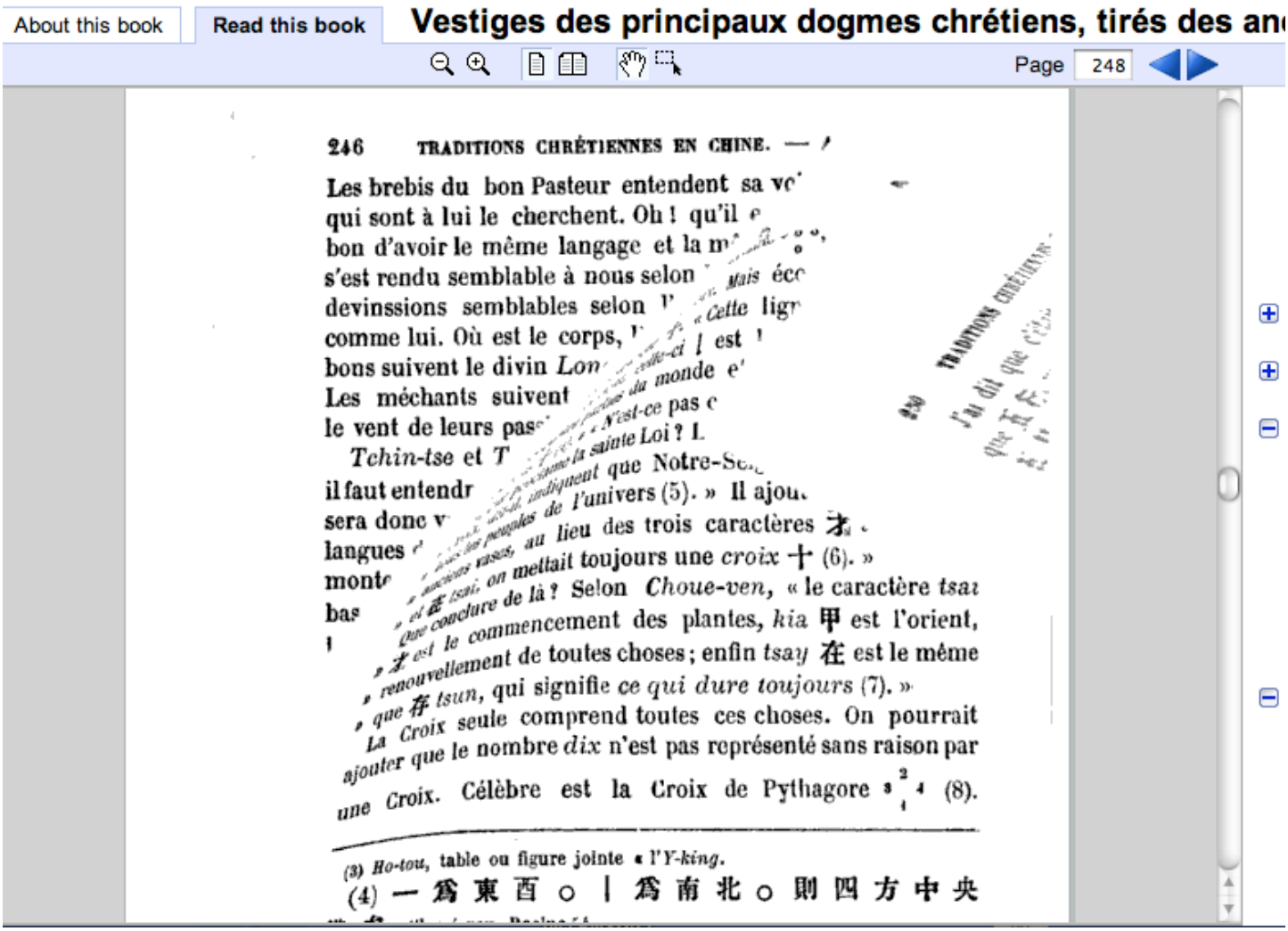


Image III

Page 249 – also lost due to page being turned during scan.

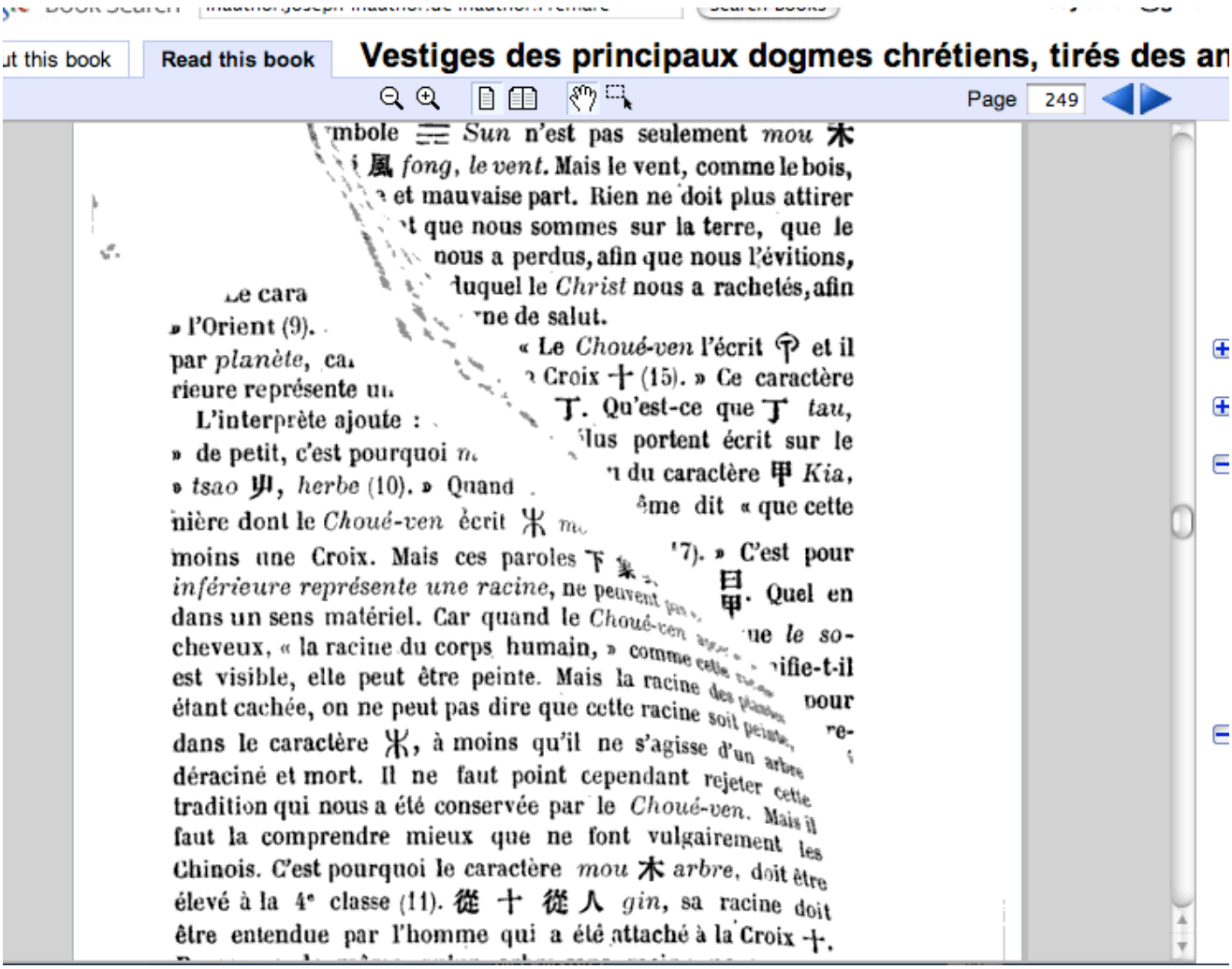


Image IV

Page 402 – listed three times, two page '402's are simply white pages scanned, as below:

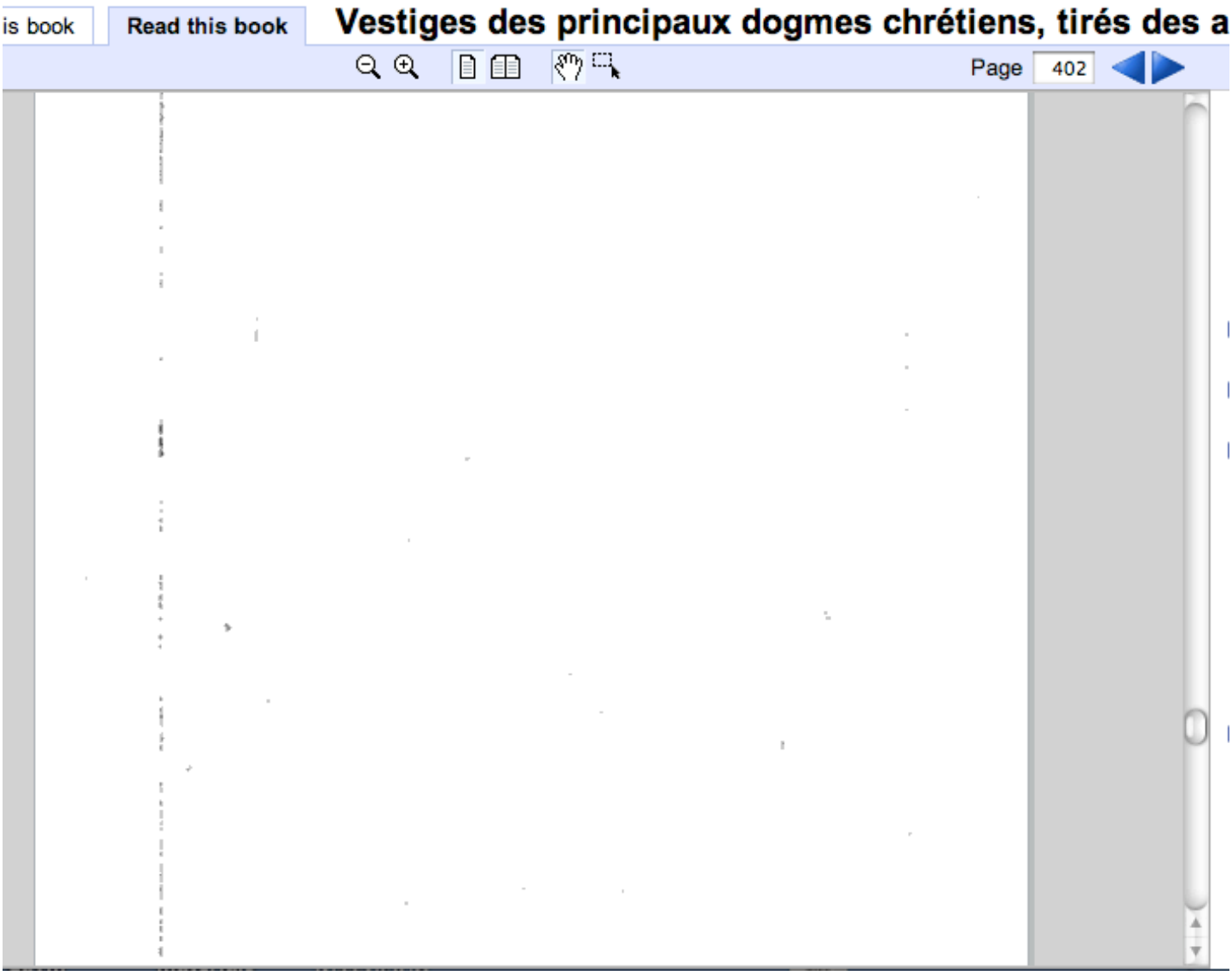
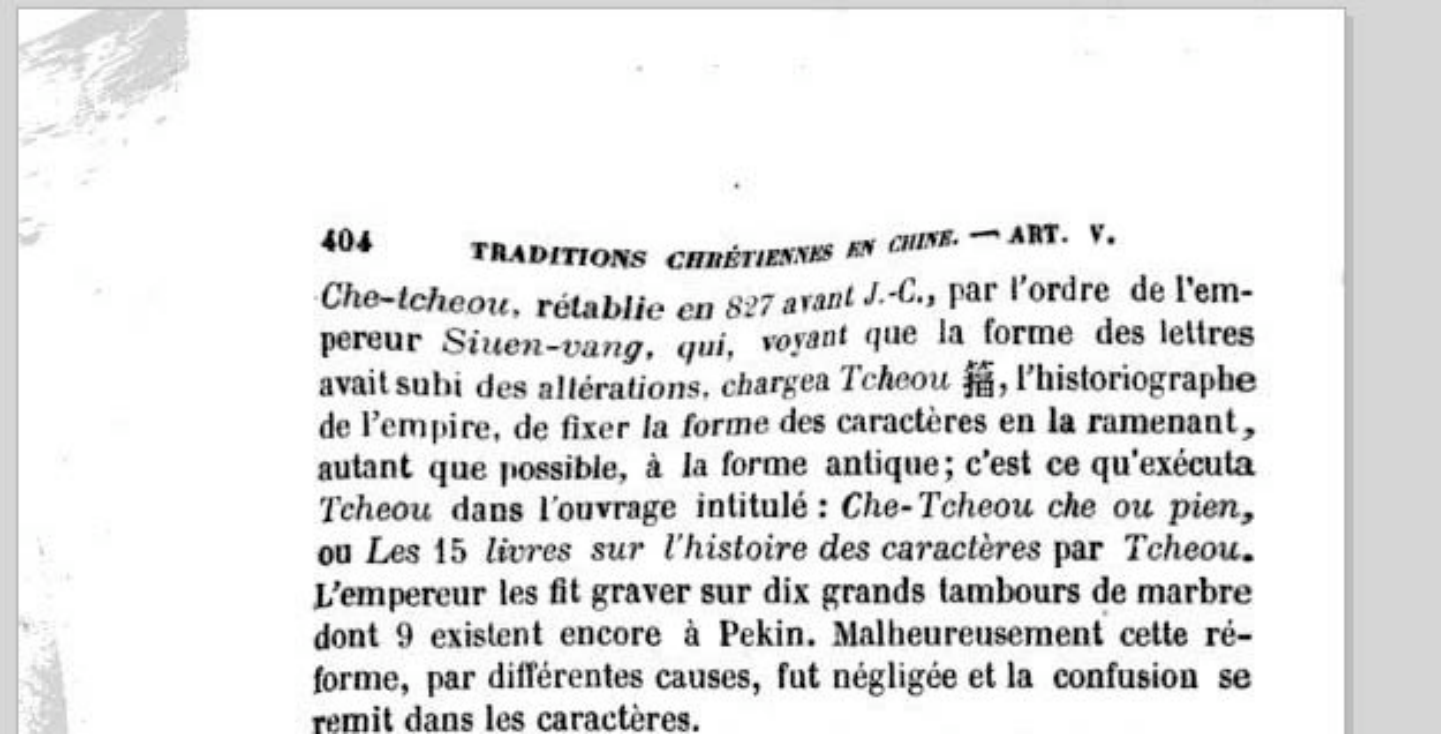


Image V

Page 404 – moved slightly during scan



Che-tcheou, rétablie *en 827 avant J.-C.*, par l'ordre de l'empereur *Siuen-vang, qui, voyant* que la forme des lettres avait subi des altérations, *chargea Tcheou* 籀, l'historiographe de l'empire, de fixer *la forme* des caractères en la ramenant, autant que possible, à la forme antique; c'est ce qu'exécuta *Tcheou* dans l'ouvrage intitulé : *Che-Tcheou che ou pien*, ou *Les* 15 *livres sur l'histoire des caractères* par *Tcheou*. L'empereur les fit graver sur dix grands tambours de marbre dont 9 existent encore à Pekin. Malheureusement cette réforme, par différentes causes, fut négligée et la confusion se remit dans les caractères.

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt	**Date of investigation**: August 23, 2008

Book title: Ensayo sobre la historia de Bolivia
Author: Manuel José Cortés
Year of Publication: 1861
Place of Publication: Sucre, Imprenta de Beeche
ISBN number?: NA
Content/Disciplinary area: History of Bolivia

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	Original provided by Harvard University
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Metadata provided accurate
MARC record available?	No
Can the book be searched through an OPAC? If so, which one?	No
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	NA

SCAN SEARCH

search paradigm settings (example: every 10 double-pages)	scanned through entire book

Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
13, 16	some extreme lightness in upper right-hand corner (may be due to state of original book)			
no other major anomalies in this book				

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Word	Comment
	(Is OCR search permitted): Y / N
revolucion / revolución	successful detection (accent mark does make a difference)
Oruro	successful detection
Yungai	successful detection
indio	successful detection
juzgamiento	successful detection
Isidoro Belzu	successful detection
i	not detected --- this letter is used often in place of ´y´, for ´and´ in Old Spanish (see page 186, 187 for examples)
Perú	successful detection
Olañeta	successful detection, except for on page 193. On this page, the word is divided between text lines. Unfortunately, when written in this way, it is not detected by OCR.
general, jeneral	jeneral is successfully detected, general no. Interesting way to see how ´general´ is spelled in this book. Researcher must consider all Old Spanish spellings when conducting search paradigms.

IMAGE ASSESSMENT	
If the original text includes graphics/illustrations, are they included in the electronic version?	No images in this book
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	
Is the scan on a gray-scale or color scale?	
Quality – does it look like the real thing?	
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	
Other notes	

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

GENERAL OBSERVATIONS	
Comment on ease of discovery in integrated interface (OPAC or better)	Easily accessible in Google Books
Overall quality of scan of this book	Good quality
Ease of access to full text	Yes
Ability to copy, quote, cite?	Yes
Ability to search?	Yes
Ability to download?	Yes
Ability to print?	Yes
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	seemingly without flaws – however, this book employs a lot of Old Spanish orthographies. Any researcher would have to take this into account -- there is even some spelling variation within the book itself.
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Absolutley
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes
Issues with this book that would affect usefulness for scholars	See above comment.
Can this book be shared with other scholars/students?	Yes, by emailing the URL
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	NA
Other notes	This is a wonderful resource providing a window into Spanish spellings and words used over a century ago

Book URL:
http://books.google.com/books?id=KV8CAAAAYAAJ&pg=PA29&dq=potosi&lr=lang_es&as_brr=1&vq=%22Ya+es+tiempo+de+organizar+un+sistema+nuevo+de+gobierno,+fundado+en+los+intereses+de+nuestra+patria%22&source=gbs_quotes#PPT1,M1
Date accessed: August 23, 2008

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt **Date of investigation**: 6/19/08 - 6/23

Searches for today:
1. Today I attempted to find a classic diary journal well known to linguists- *Speech Development of a Bilingual Child – a Linguist's Record*. Written by Werner Leopold (and in 4 volumes), this is a class diary case (1939-1949) describing his daughter's simultaneous acquisition of German and English. Unfortunately this is not available with Google Books or in GEORGE. Northwestern U (publisher) also does not have electronic copy.
2. Bickerton, Derek. (1975). *Dynamics of a Creole System*. London: CUP. Not available on Google Books.
3. Carmen Muñoz is a well-known linguist from Spain. Neither of her two books were available for preview on Google.
4. I did a search for books written in Portuguese (in Google Books), with the subject designated as 'gramatica' (language set to Portuguese). I wanted to find A book on grammar/linguistics in Portuguese! Google gave me 1920 results – I had to go through 18 pages of search results to find a book with full preview available (most were no preview, few had snippet views). It is called 'Gramática Portugueza, written by Manoel Diaz de Souza in 1804. I'll do this text. **Note- Google incorrectly spells the result hit as Grammática Portugueza, with 2 m's.**

Book title: Gramática Portugueza
according to WorldCat, the official title is "Gramatica portugueza ordenada segundo a doutrina dos mais celebres gramaticos conhecidos, assim nacionaes como estrangeiros, para facilitar á mocidade portugueza o estudo de lêr e escrevêr a sua propria ligua, e a intelegencia das outras em que se quizer instruir"
Author: Manoel Diaz de Souza
Year of Publication: 1804
Place of Publication: Coimbra, Na Real Imprensa da Universidade (Portugal) – **this publication information is not given in Google Books- I had to obtain it through WorldCat**
ISBN number?:
Content/Disciplinary area: Portuguese Grammar

AVAILABILITY SEARCH

Source of scanned book:	Google Books. They scanned entire book – even front leather cover, which is in color. This book was scanned on August 11, 2006 – "it is an original from Oxford University" (according to Google)
Comments on condition/provenance of book:	some scanning quality very light and poor visibility- may be due to age of book?
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Very easy. Metadata NA.
MARC record available?	No
Can the book be searched through an OPAC? If so, which one?	I could not find it in George. WorldCat has this reference however. Google Books is only place where researchers can find full copy.
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	NA. Table of contents, other metadata not hyperlinked.

SCAN SEARCH

search paradigm settings (example: every 10 double-pages) Scanned through entire book.

Page(s):	readability/ quality of scan	anomalies?	other?	other?
xviii	poor	spine of book not opened all the way – many of the page	much of the text is very light on this page- a few searches, such as *imitar*	

		numbers in the index are cut off on this page	or *uzão* are not detected	
1	some parts of text too light to be detected	Most pages have text in lower left corner—this page's word is *todos* and is not detected by search browser		
2	text color (scan) is good	spine of book not opened all the way – much of text is cut off	(see image 1)	
3, 4, 6, 11, 26, 40, 48, 96, 111, 112, 121, 123, 124, 130, 145 (incorrectly, recognized as 147), 150, 152, 171, 193, 195, 196, 198, 210, 215, 217, 219, 236, 239, 244, 265, 266, 275	some scan blurriness of words/letters close to spine of book and words have been cut off	(pages 4, 171, 244 are particularly poorly scanned by the spine, with some words being cut in half)	(page 42 is cut off RIGHT at beginning of word – a printed out version would prove to be difficult to bind for this reason.	246-247's top corners are blurred and stretched due to book being moved during scan
6	footnote on page 6 very clear			
7		On this page, the author distinguishes between vowels – os agudos, os medios, os graves. Does so as such: á é í ó ú ; â ê î ô û ; à è ì ò ù . Broswer does not detect these/ and or confuses the accent marks. Search result specifically gives me: á é i ú o ; â ê í ô u ; à è i ò u. While most of these vowels **within words** are detected properly, the misreading on page 7 of the vowels as single entities is a significant discrepancy in a Portuguese grammar book.		
16	overall good, some scanned text is too light.	Blurriness on end of one word - tambem?		
29		some words have text that is so whited out it prevents them from being detected by searches. (ex. si-labas)		
31		this page has particularly high amount of very light text – hard to read and detect		
32-39, 51-95, 100-105	scan is quite good quality			
106, 182-183	these pages have heavy	This may be due to	see image 2	

	print marks of text from another page, though it does not impede readability.	original text		
126-127!	These pages are COMPLETE scan anamolies!	Completely illegible – page may have been moved during scan?	What appears to be finger is present on page 127.	See image 3.
142-145, 147	These pages are missing entirely from scan.	Note - After page 146, page 145 is present. However, Google Books page browser lists this as page 147. Starting here, the book pages become misaligned.	order is 141, 146, 145 (listed as 147), and then 148 – here ordering is fixed.	
179-180	blurriness, scan quality quite poor on these pages			
202		folio marker at bottom of the page has been cut off		
216	scan quality of this page very poor	see image 4		
222-223	these pages are missing entirely from scan.	After page 220, Google book page browser recognizes page 221 as 223.	Page 224 is correctly numbered, both in scan and by Google Books.	
224	This page was turned during scan – page contents are at an angle			
227	some text in footnote on page 227 has been cut off			
248-251	these pages are missing entirely from scan.	Google Books incorrectly recognizes page 247 as 251.	On page 252 pages are realigned with Google Books.	
255	This page has significant blurriness/ text cutoff	see image 5		
262-263	these pages are missing entirely from scan.	page 264 is incorrectly recognized by Google as page 262.	Google Books then jumps to page 252, where GB's and scan pages are realigned	

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	Comment (Is OCR search permitted): Y / N
proemio	This word is detected sometimes; sometimes it is not. For example, it is a subtitle on page 3 and is not detected by search.
primeira / PRIMEIRA	Same as above—this word (capitalized) is a subtitle on page 3 and is not detected – in caps or lower case.
expressão º	Same as above – this word detected half of the time. (not detected on page 3). Due to space limitation on the folios, many early authors divided sentences with this symbol – a small circle in upper half of text space. Not only did Google Books search browser NOT recognize this symbol, it mistakenly confused it with erroneous specks. For example, page 10 in this text is black, but there are some black specks from the scan (or perhaps age of folios). The search browser suggests this symbol might be on that page, among others. Does not recognize symbols as this one.

etimologia	Search browser finds this word successfully.
myſterio / ſ	Several old Portuguese and old Spanish texts used this symbol, ſ, to indicate a voiceless alveolar fricative. Search browser does not detect this symbol. ſegundo on page 43 is an example.
vogaes	Successfully detected, on pages 7, 8, 15, but not detected on page 6 for some reason.
criança, braço	Did this search to test letter ç. Search detected successfully.
consoantes	successfully detected.
Psalterio	Not detected. This word is italicized in the text – possible reason for it not being detected?
Assim	Improperly detected – unsuccessful on page 29, perhaps because first two letters are capitalized as the word starts the paragraph. Detected successfully on pages 31, 33, 35, 36, etc.
regra	detected successfully

IMAGE ASSESSMENT

If the original text includes graphics/illustrations, are they included in the electronic version?	**NA – no images in this text**
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	
Is the scan on a gray-scale or color scale?	
Quality – does it look like the real thing?	
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	
Other notes	

GENERAL OBSERVATIONS

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Comment on ease of discovery in integrated interface (OPAC or better)	Easily accessible through Google Books. Could not find it in an OPAC.
Overall quality of scan of this book	Mediocre. Blurriness, text ink merging, whiteness may be due to age of folios of original text.
Ease of access to full text	Very
Ability to copy, quote, cite?	No. For example, the way in which this author defines a verb is pretty neat/humorous/eloquent on page 49. I'd like to highlight the text, copy, and paste it into a word document - Google Books does not permit.
Ability to search?	Yes
Ability to download?	Yes
Ability to print?	Yes- if one downloads original doc
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	Some major issues with symbols, Portuguese accents, old and decorative letters, most foliation symbols or letters are not detected, blurred letters, words in spine of book.
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	No, given that several pages are missing from this text. Also, some text is cut out from book spine.
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	No.
Issues with this book that would affect usefulness for scholars	This book is a historical grammar book of the Portuguese language. It's description of the language dates back to when, for example, Portuguese still used vós forms and does not include the pronoun você. However, search paradigms are sometimes impossible due to Google Book's inability to detect certain symbols (example ∫ symbol). Also, spine of book was not opened all of the way on many pages and therefore this book's scan leaves out/ cuts out some text. Several Portuguese accents are not detected or are improperly detected. Lastly, many of the foliation words or letters are either not detected at all or are cut out. Old texts were sometimes printed on folios that were folded four-way. Providing a heading or a foliation word – typically the first word of the next page – was the way for the editor to keep the pages aligned. Foliation words are usually below the text line. For some reason, foliations in this book are almost always NOT detected by Google Books. This is odd given that for the most part, their scan quality is good. For example, **actiua** on page 120 is not detected (also tried activa – still did not work). **APEN** on page 214 is not detected.
Can this book be shared with other scholars/students?	Yes- via emailing book URL
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	NA
Other notes	This book has an elaborate symbol for the letter Q. I wanted to do searches for it but could not – the symbol is not available in Microsoft word images. See page 19 for example of the Q.

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Book URL: http://books.google.com/books?id=QYYCAAAAQAAJ&pg=PA1&dq=gramatica&lr=lang_pt&as_brr=0#PPA1,M1

Date accessed: 6/19, 6/22, 6/23

Image 1 – example where text is cut off from spine of book not being opened all of the way (right side)

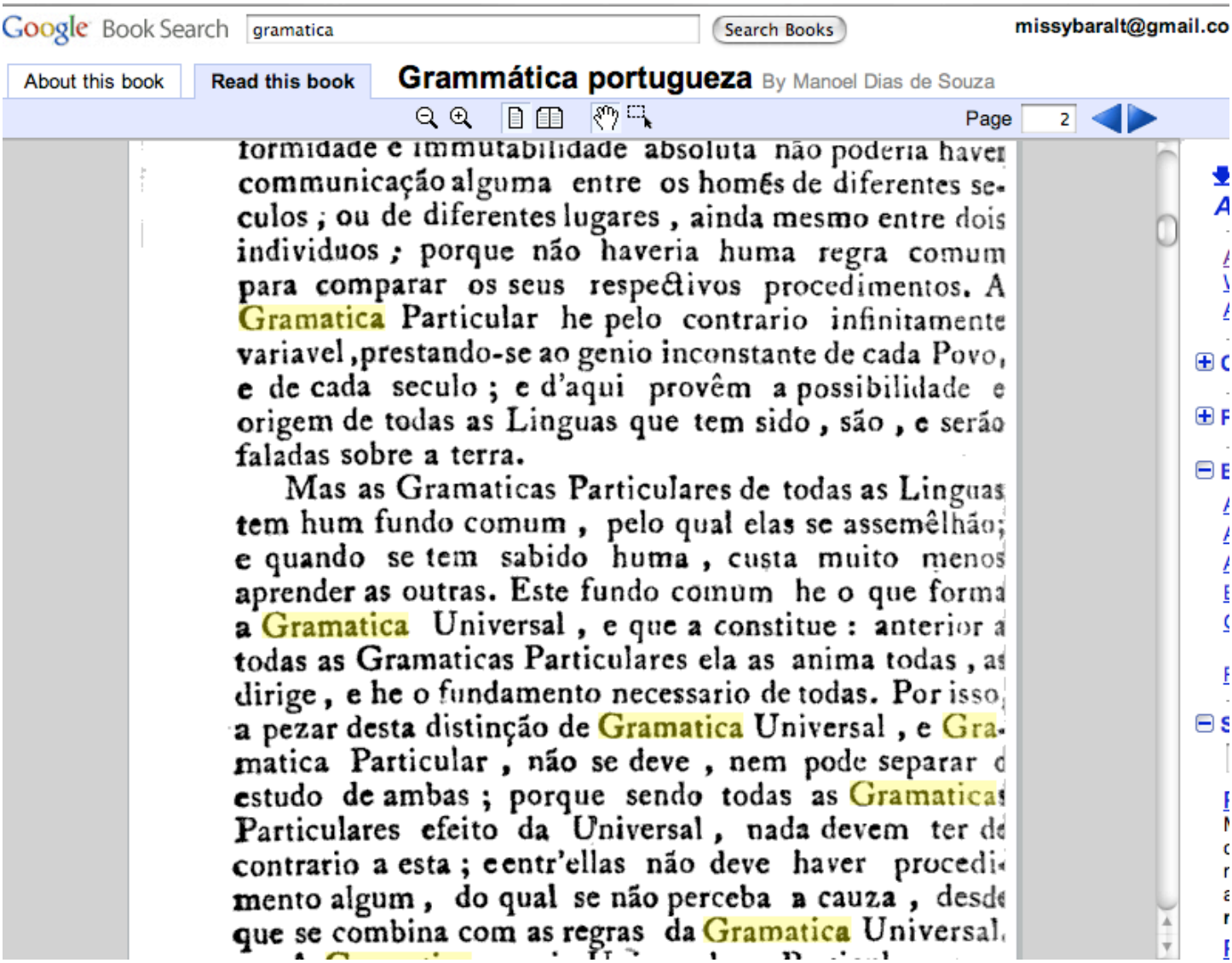
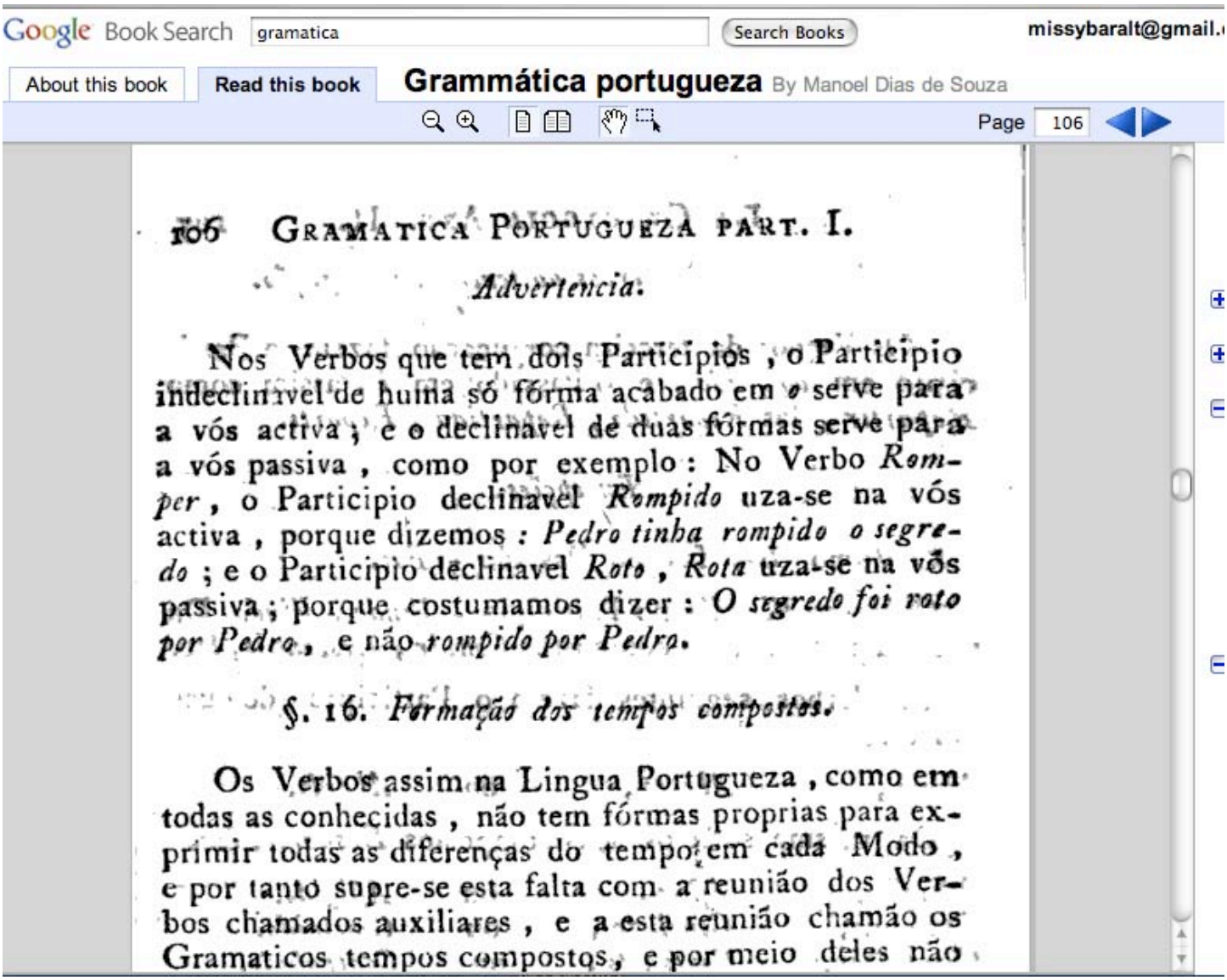


Image 2
Page 106 with text imprints from another page (this is probably identical to the original and is thus more of an issue with old texts)



106 GRAMATICA PORTUGUEZA PART. I.

Advertencia.

Nos Verbos que tem dois Participios, o Participio indeclinavel de huma só fórma acabado em *o* serve para a vós activa; e o declinavel de duas fórmas serve para a vós passiva, como por exemplo: No Verbo *Romper*, o Participio declinavel *Rompido* uza-se na vós activa, porque dizemos: *Pedro tinha rompido o segredo*; e o Participio declinavel *Roto*, *Rota* uza-se na vós passiva; porque costumamos dizer: *O segredo foi roto por Pedro*, e não *rompido por Pedro*.

§. 16. *Formação dos tempos compostos.*

Os Verbos assim na Lingua Portugueza, como em todas as conhecidas, não tem fórmas proprias para exprimir todas as diferenças do tempo em cada Modo, e por tanto supre-se esta falta com a reunião dos Verbos chamados auxiliares, e a esta reunião chamão os Gramaticos tempos compostos, e por meio deles não

Image 3
Page 127 – Complete scan anomaly. This page is thus lost from overall scan job.

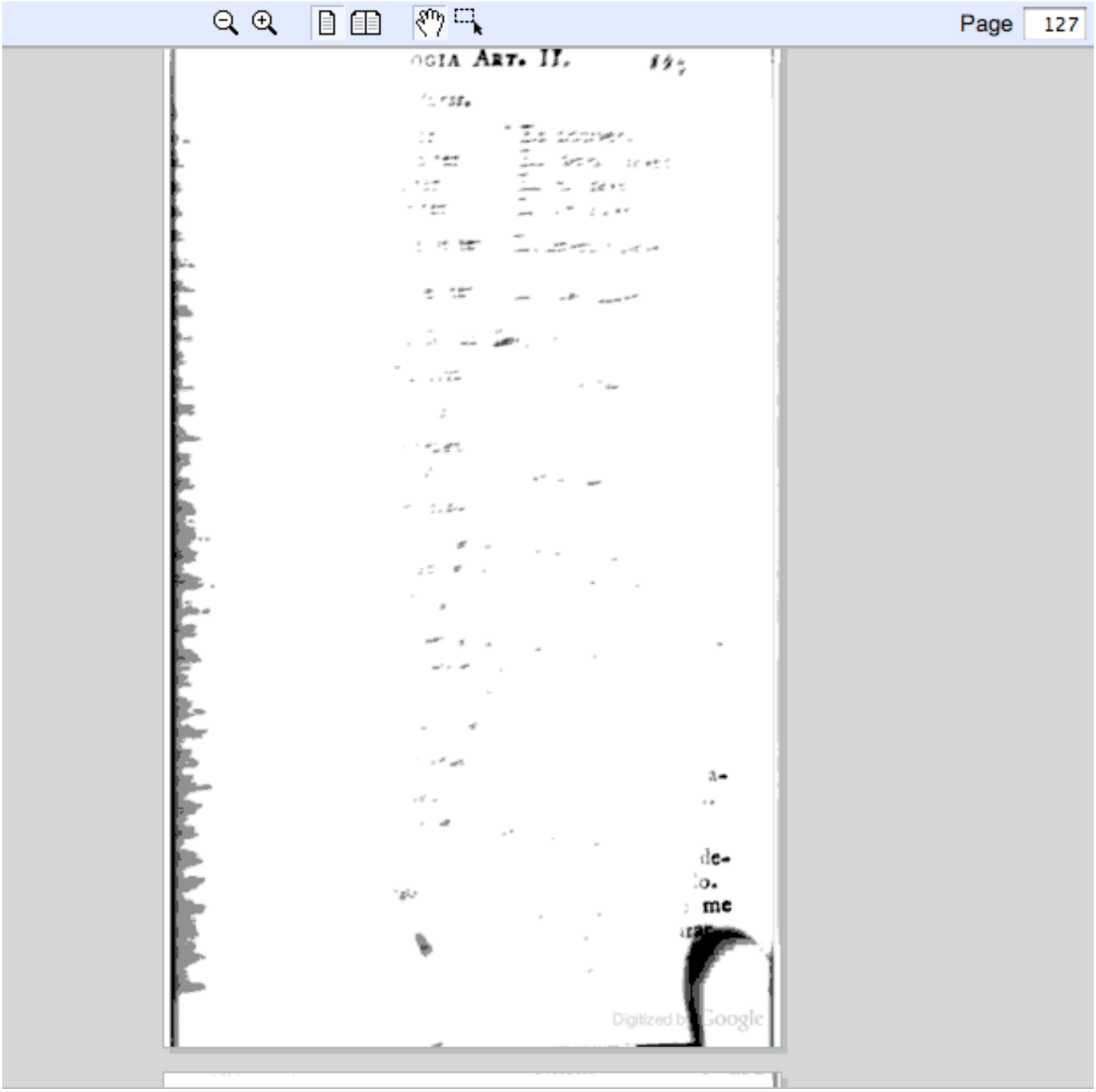
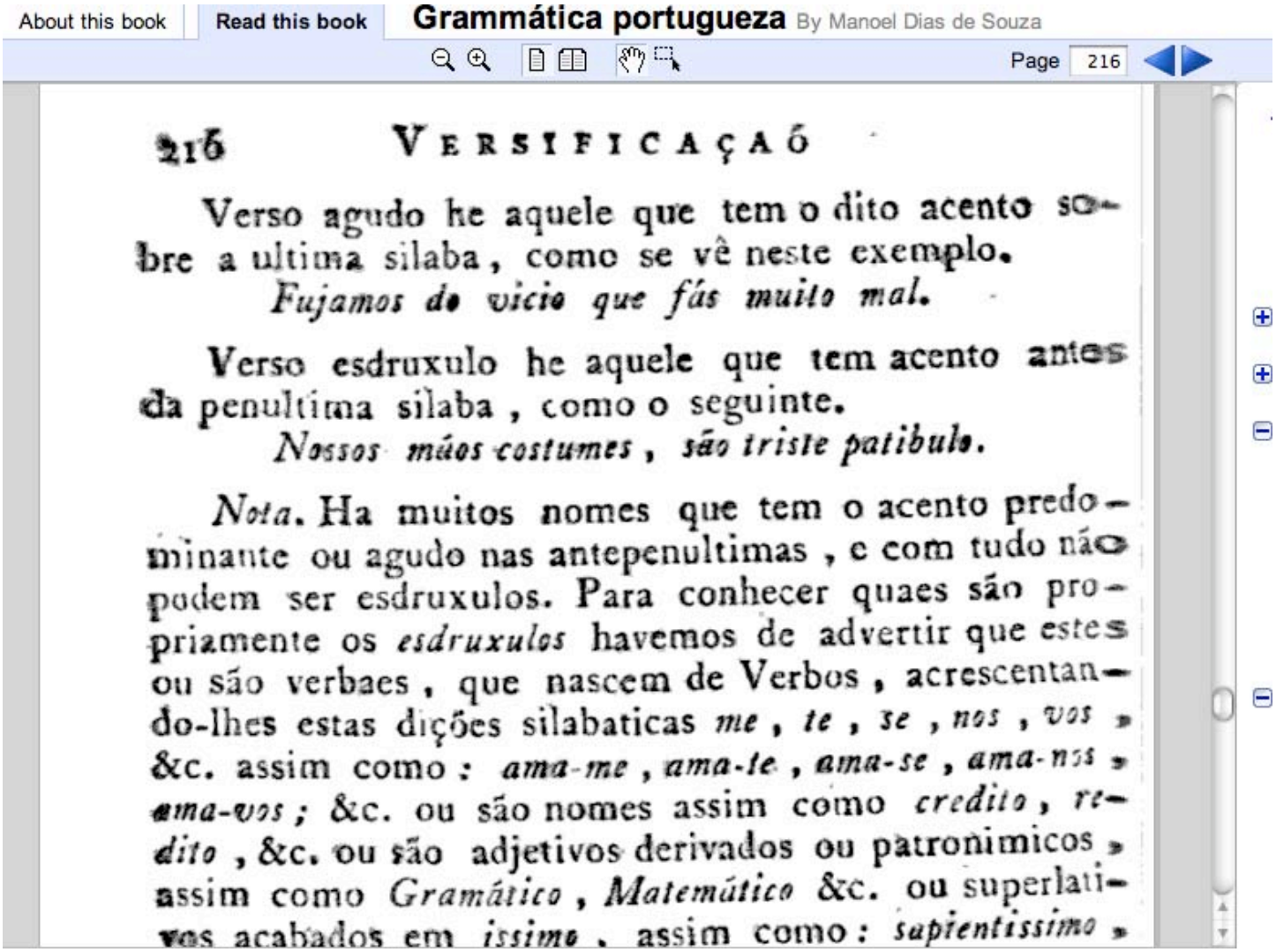


Image 4
Page 216 – note blurriness, stretched letters from scan



216 VERSIFICAÇAÓ

Verso agudo he aquele que tem o dito acento sobre a ultima silaba, como se vê neste exemplo.

Fujamos do vicio que fás muito mal.

Verso esdruxulo he aquele que tem acento antes da penultima silaba, como o seguinte.

Nossos máos costumes, são triste patibulo.

Nota. Ha muitos nomes que tem o acento predominante ou agudo nas antepenultimas, e com tudo não podem ser esdruxulos. Para conhecer quaes são propriamente os *esdruxulos* havemos de advertir que estes ou são verbaes, que nascem de Verbos, acrescentando-lhes estas dições silabaticas *me*, *te*, *se*, *nos*, *vos*, &c. assim como: *ama-me*, *ama-te*, *ama-se*, *ama-nos*, *ama-vos*; &c. ou são nomes assim como *credito*, *redito*, &c. ou são adjetivos derivados ou patronimicos, assim como *Gramático*, *Matemático* &c. ou superlativos acabados em *issimo*, assim como: *sapientissimo*,

Image 5

DA ORTOGRAFIA ART. I. 255

Verbos	Nomes	Verbos	Nomes
Estímulo	Estímulo	Prognostíco	Prognóstico
Fóra	Fóra	Publíco	Público
Fôrça	Fôrça	Perpetúo	Perpétuo
Fôrma	Fôrma	Régo	Rêgo
Fôrra	Fôrra	Reforço	Reforço
Fôrro	Fôrro	Rógo	Rôgo
Gêlo	Gêlo	Regúlo	Régulo
Gôsto	Gôsto	Refrésco	Refrêsco
Gôzo	Gôzo	Róla	Rôla
Govêrno	Govêrno	Ródo	Rôdo
Grêlo	Grêlo	Renuncía	Renuncia
Grêta	Grêta	Sóbre	Sôbre
Glória	Glória	Sóldo	Sôldo
Hábito	Hábito	Socégo	Socêgo
Injúria	Injúria	Séco	Sêco
Incômodo	Incômodo	Séca	Sêca
I'ntimo	I'ntimo	Sélo	Sêlo

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt **Date of investigation**: Sunday, August 17, 2008

Book title: Guatemala: The Land of the Quetzal
Author: William Tufts Brigham
Year of Publication: 1887
Place of Publication: Scribner, London
ISBN number?: NA
Content/Disciplinary area: history, Guatemala

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	Seemingly good condition, original was provided by Harvard University
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Easily accessible on Google books. Metadata provided is good- includes content subjects and other editions of this book.
MARC record available?	NA
Can the book be searched through an OPAC? If so, which one?	No
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	Very helpful. Table of contents is hyperlinked so that the scholar can jump to each respective section.

SCAN SEARCH

search paradigm settings (example: every 10 double-pages) Searched through entire book

Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
108			scan of pictures is at erroneous angle – could be fixed for final version online	
No other anomalies in this book!				

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	**Comment** (Is OCR search permitted): Y / N
Livingston	successful recognition
quetzal	successful recognition
Lisandro Barillas	successful recognition
ritrato	successful recognition

trapiche	result shows hit on page 452, but this word also appears on page 196
Ætna	unsuccessful – search yielded no results, even though this word appears on page 198
Quichés	unsuccessful; this word appears on page 235 but is not detected
Xmucane	successful recognition
Yaxcanulmucamob	successful recognition
Sipacua	successful recognition

IMAGE ASSESSMENT

If the original text includes graphics/illustrations, are they included in the electronic version?	Yes, to include some real photos. See Figure 1 and below. This book also includes a list of all images, which have been hyperlinked by Google. This is quite helpful – a feature I have not seen yet in all of my searches. From an Ancient Manuscript.
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	NA
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	PDF
Is the scan on a gray-scale or color scale?	gray-scale
Quality – does it look like the real thing?	Yes – wonderful images!
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	Absolutely – nice that this resource is available online
Are the illustrations searchable?	Only by their title- all of the images/pictures in this text have titles, and I can search for them that way
Ability to print illustrations? (is there a watermark?)	Yes - no watermark
Are the images downloadable? (file size, pixel quality)	Only within the entire text, not as separate images
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	I don't think so
Other notes	

GENERAL OBSERVATIONS

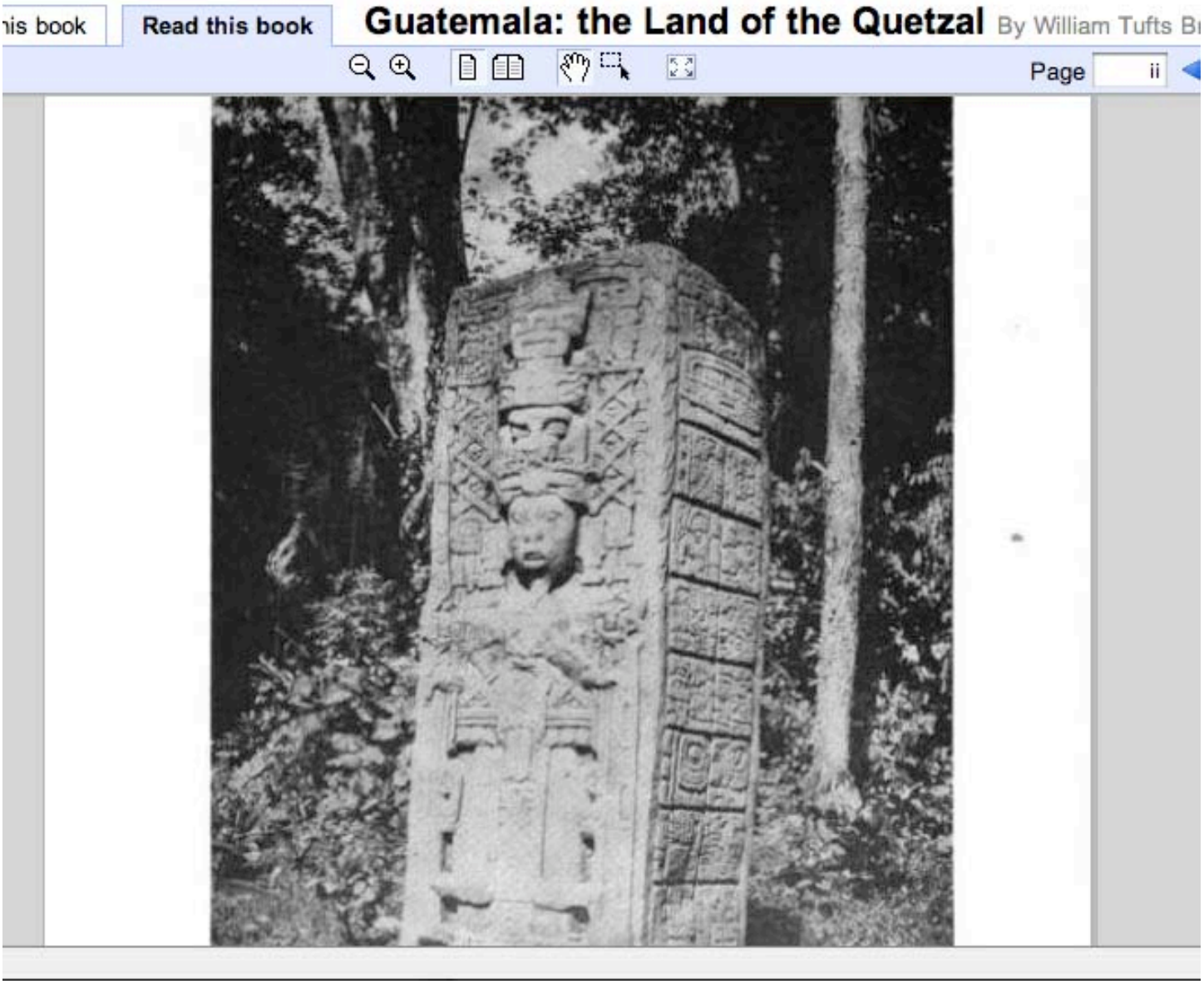
CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Comment on ease of discovery in integrated interface (OPAC or better)	Easily accessible in Google Books
Overall quality of scan of this book	excellent, hardly any anomalies
Ease of access to full text	100% - full view
Ability to copy, quote, cite?	Yes
Ability to search?	Yes
Ability to download?	Yes
Ability to print?	Yes
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	some words are detectable, but others are not
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Absolutely
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes
Issues with this book that would affect usefulness for scholars	Some words, for example those containing accent marks, are not detected by an OCR search for some reason
Can this book be shared with other scholars/students?	Yes, by emailing the URL
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	NA
Other notes	

Book URL:
http://books.google.com/books?id=8EITAAAAYAAJ&pg=PA440&dq=adquisicion+de+lengua&lr=&as_brr=1#PPT1,M1

Date accessed: August 17, 2007

Figure 1, example of picture in book



CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt **Date of investigation**: Saturday, April 9, 2008

Book title: Handbook of Second Language Acquisition
Author: Catherine Doughty and Michael H. Long
Year of Publication: 2003
Place of Publication: Blackwell Publishing
ISBN number?: 0631217541
Content/Disciplinary area: linguistics, second language acquisition. This book is considered a critical guide for any scholar of SLA.

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	Limited preview by Google Books, however this book has detailed metadata provided and hyperlinked chapters for scholars to view. (See Figure 1)
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Wonderful – great amount of detail is given in metadata.
MARC record available?	na
Can the book be searched through an OPAC? If so, which one?	No, however, Georgetown's GEORGE system provides a quick view, including view of cover image, table of contents, summary of book, and even a book review.
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	Very detailed.

SCAN SEARCH

search paradigm settings (example: every 10 double-pages) searched through entire book

Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
2, 18, 75-78, 84-510, 515-537, all pages after 540			These pages are not included in this book for view on Google Books. I was surprised to see that almost the entire first 3 chapters are viewable (with reference sections!) but then after page 84, Google does not allow viewing	
	NO ANOMALIES – state of this book (what is viewable) is in perfect condition.			

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	Comment (Is OCR search permitted): Y / N
functionism	successful recognition – all with yellow highlighting!
connectionism	successful recognition
behaviorism	successful recognition
fossilization	successful recognition
SLA	successful recognition
UG	successful recognition
interlanguage	successful recognition
nativism	successful recognition
constructions	successful recognition
chunking	successful recognition

IMAGE ASSESSMENT

If the original text includes graphics/illustrations, are they included in the electronic version?	No images in this text
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	
Is the scan on a gray-scale or color scale?	
Quality – does it look like the real thing?	
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	
Other notes	

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

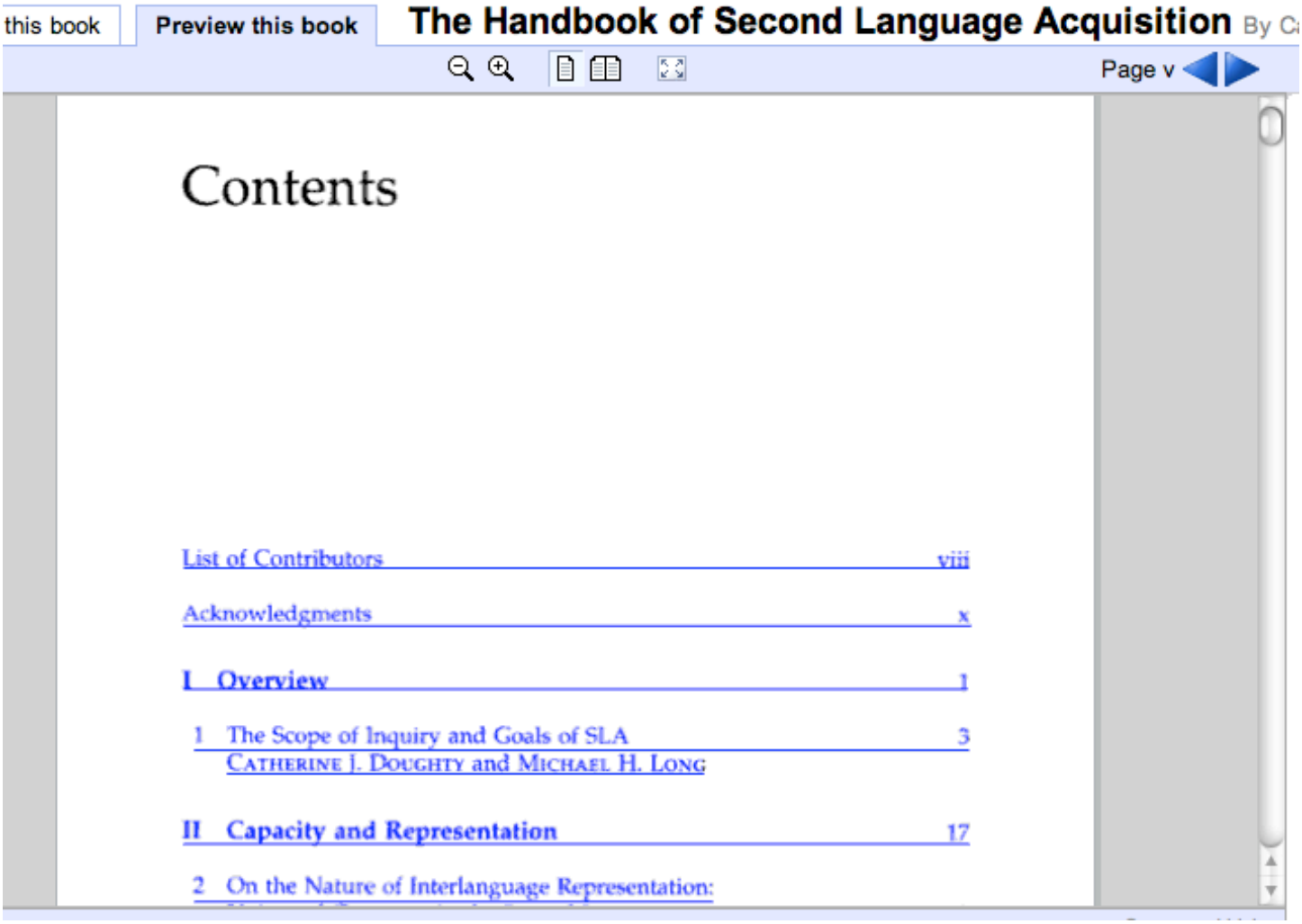
GENERAL OBSERVATIONS

Comment on ease of discovery in integrated interface (OPAC or better)	Easily accessible in Google Books, however, this book is limited preview only. The scholar can read up to about chapter 4 and then all of the pages are excluded.
Overall quality of scan of this book	Excellent- I'd say perfect
Ease of access to full text	Except those pages excluded from view
Ability to copy, quote, cite?	No
Ability to search?	Yes
Ability to download?	No
Ability to print?	No
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	100% perfect
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Absolutely
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes
Issues with this book that would affect usefulness for scholars	Only that it is not all available for preview to scholars
Can this book be shared with other scholars/students?	By emailing the link, yes
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	No
Other notes	Important resource- many scholars wish they'd make this fully available online

Book URL: http://books.google.com/books?id=tH5-z_zgtQUC&printsec=frontcover&dq=inauthor:alison+inauthor:mackey&lr=&as_brr=0&source=gbs_summary_r&cad=0

Date accessed: Saturday, August 9, 2008

Figure 1 – example of hyperlinked chapters and book sections



CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt	**Date of investigation**: August 20, 2008

Book title: How Language Works: How Babies Babble, Words Change Meaning, and Languages Live or Die
Author: David Crystal
Year of Publication: 2007
Place of Publication: Avery
ISBN number?: 158333291X
Content/Disciplinary area: language, language acquisition

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	Seemingly good quality, this book is a limited preview due to copyright issues
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Metadata includes content and some pictures that are in the book
MARC record available?	NA
Can the book be searched through an OPAC? If so, which one?	No
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	Accurate and detailed

SCAN SEARCH

search paradigm settings (example: every 10 double-pages)	searched through entire book

Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
19 – 122, 126 – 131, 144 – 359, 361, 364 – 379, 381, 383, 385 – 388, 391, all pages after 393	these pages are not included in the preview			
	no anomalies in this book			

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	**Comment**
	(Is OCR search permitted): Y / N
acquisition	successful recognition
Austronesian	successful recognition

language death	successful recognition
Uralic	successful recognition
dyslexia	successful recognition
linguistic skills	successful recognition
learning to write	successful recognition
decode	successful recognition
sign languages	successful recognition
paralinguistic	successful recognition

IMAGE ASSESSMENT	
If the original text includes graphics/illustrations, are they included in the electronic version?	No images available in preview pages
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	
Is the scan on a gray-scale or color scale?	
Quality – does it look like the real thing?	
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	
Other notes	

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

GENERAL OBSERVATIONS	
Comment on ease of discovery in integrated interface (OPAC or better)	Easily accessible in Google books, but unfortunately not all pages can be viewed due to copyright issues
Overall quality of scan of this book	excellent!
Ease of access to full text	No
Ability to copy, quote, cite?	No
Ability to search?	Yes
Ability to download?	No
Ability to print?	No
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	seemingly perfect
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Yes (but must include all pages)
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes
Issues with this book that would affect usefulness for scholars	while searching can be done, limited preview will not sure the author the page or context in which the search is found
Can this book be shared with other scholars/students?	Yes, by emailing the URL
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	NA
Other notes	

Book URL:
http://books.google.com/books?id=t7FSSgBqaEgC&pg=PT377&dq=how+language+work+how+babies+babble&sig=ACfU3U2O22ecj3qucoDqsaj3l9ik6C6bWg#PPT26,M1

Date accessed: August 20

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt **Date of investigation**: August 16, 2008

Book title: La psychologie de l'enfant: l'art et la poésie chez l'enfant
Author: Bernard Perez
Year of Publication: 1888
Place of Publication: F. Alcan, Paris
ISBN number?: NA
Content/Disciplinary area: child psychology and development

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	Seeming good- original digitized from New York Public Libraries
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Easily accessible in Google Books. Metadata provided includes content by subject and popular passages.
MARC record available?	NA
Can the book be searched through an OPAC? If so, which one?	NA
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	

SCAN SEARCH

search paradigm settings (example: every 10 double-pages) Scrolled through entire book as not to miss anything

Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
43		This page has some miscoloration (almost as if scan was colorized- i.e. scanner's finger shows in color). Does not impede readability however		
74-75		Scanner's fingertips show		
	No other anomalies!			

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	**Comment**
	(Is OCR search permitted): Y / N
Langage, langue	Both successful hits. Yellow highlighter tool worked as well, even in small footnotes
l'acquisition du langage	Successful recognition

Conscience	Successful recognition
Remarquer	Successful recognition
d'entrée	Successful recognition
Production	No hits
Interaction	No hits
Jouer	Successful recognition
Pair	No hits
adulte	Successful recognition

IMAGE ASSESSMENT

If the original text includes graphics/illustrations, are they included in the electronic version?	No images in this text
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	
Is the scan on a gray-scale or color scale?	
Quality – does it look like the real thing?	
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	
Other notes	

GENERAL OBSERVATIONS

Comment on ease of discovery in integrated interface (OPAC or better) Overall quality of scan of this book	Excellent quality in this book by Google Books, very easily accessible
Ease of access to full text	Yes

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Ability to copy, quote, cite?	Yes
Ability to search?	Yes – no recognition problems with this books
Ability to download?	Yes
Ability to print?	Yes, if you download original document
Scanned book allow for consulting of contents?	Yes – excellent scanability
How successful is OCR recognition?	100% accurate
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Yes, though I'd eliminate the few instances of scanner fingers present in scan
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes
Issues with this book that would affect usefulness for scholars	None, pagination system for this book is perfect, excellent searchability
Can this book be shared with other scholars/students?	Yes, by emailing the URL
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	NA
Other notes	Great quality

Book URL:
http://books.google.com/books?id=Pa8NAAAAYAAJ&printsec=frontcover&dq=subject:%22child+psychology%22+date:1500-1900&lr=&as_brr=0#PPA104,M1

Date accessed: Saturday, August 16

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt	**Date of investigation**: August 23, 2008

Book title: La Vida Imperial de Potosí
Author: Julio Lucas Jaimes
Year of Publication: 1905
Place of Publication: Talleres gráficos de L. J. Rosso, Buenos Aires
ISBN number?: NA
Content/Disciplinary area: town of Potosí, history

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	Original book was provided by Harvard University.
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Easily accessible in Google books for full preview. Includes accurate author and bibliographic information, content, popular passages, and pages with pictures. Google even provides the researcher which books quote these popular passages – one was quoted by 102 different books!
MARC record available?	NA
Can the book be searched through an OPAC? If so, which one?	Yes
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	Pages 537 and 538 (the index) have nice hyperlinks for the researcher to jump to those specific sections

SCAN SEARCH

search paradigm settings (example: every 10 double-pages)	scanned through entire book as to not miss anything

Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
131,	scanners fingers – in color – appear on this page! (see figure 3). Cover text			
218, 220, 240, 284,	these pages were moved during scan – ½ is illegible due to blurriness			
346	This page is missing from the book. In its place, a second scan of page 344 is present with the scanners fingers in color and in complete view			
347	This page is present THREE times in the scan. First is normal, second is a complete	Why did Google not just eliminate latter two pages from scan?		

	anomaly (so blurred illegible) and the third has scanners fingers).			
348, 406		scanner's entire hand visible in scan, in color		
354, 356	blurriness in top right corner, illegible			

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	Comment
	(Is OCR search permitted): Y / N
Viila imperial	successful detection
inca	successful detection
Tahuacco	successful detection
Cuzco	successful detection
Huaina Capac	successful detection
1545	successful detection
cuidábanse	successful detection (interesting to see morpheme placement here of reflexive pronoun)
Yurac Quilla	successful detection
BAÑOS	not detected – similar to ´laguna´ above, this word is a picture title on page 52. I presume this is the only way researchers can search for the pictures in this text, but picture text is not detected for some reason. (Figure 2)
mística	successful detection

IMAGE ASSESSMENT

If the original text includes graphics/illustrations, are they included in the electronic version?	Yes – some wonderful drawings and pictures
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	No
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	OCR. OCR recognition NA
Is the scan on a gray-scale or color scale?	gray-scale
Quality – does it look like the real thing?	Few. Most of the pictures in this book are pretty low in quality, though this may be due to the original state and age of the book scanned.
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	Absolutely
Are the illustrations searchable?	Only by subtitles, but some of these are not detected in searches. For example, the

	picture of the Tarapava Laguna on page 9: I typed in Laguna, and it was not detected on this page.
Ability to print illustrations? (is there a watermark?)	Yes – no watermark
Are the images downloadable? (file size, pixel quality)	Yes, only within entire document
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	zooming in
Other notes	See Figure 1 for an example

GENERAL OBSERVATIONS	
Comment on ease of discovery in integrated interface (OPAC or better)	Easily accessible in Google books
Overall quality of scan of this book	Average – the fact that the scanner's hand appears in five separate pages renders this scan problematic
Ease of access to full text	Yes
Ability to copy, quote, cite?	Yes
Ability to search?	Yes
Ability to download?	Yes
Ability to print?	Yes
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	somewhat – smaller text is sometimes undetectable, for example picture titles
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	At this point (with so many anomalies, i.e. scanners hand, blurriness, repetition of pages), no. If those issues were to be corrected, yes.
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	see above.
Issues with this book that would affect usefulness for scholars	Some smaller words are not detected in OCR scan. The pictures in this book are wonderful, but some are a little blurry.
Can this book be shared with other scholars/students?	Yes, by emailing the URL
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e.	NA

volumes in a set, standard words, etc.)	
Other notes	excellent quotes in Old Spanish in this book

Book URL:
http://books.google.com/books?id=D2QCAAAAYAAJ&printsec=frontcover&dq=potosi&lr=lang_es&as_brr=1#PPA63,M1
Date accessed: August 23

Figure 1

Figure 2 picture on page 52

Figure 3 : anomaly on page 131

Figure 4 page 346 (actually page 344 – page 346 is missing)

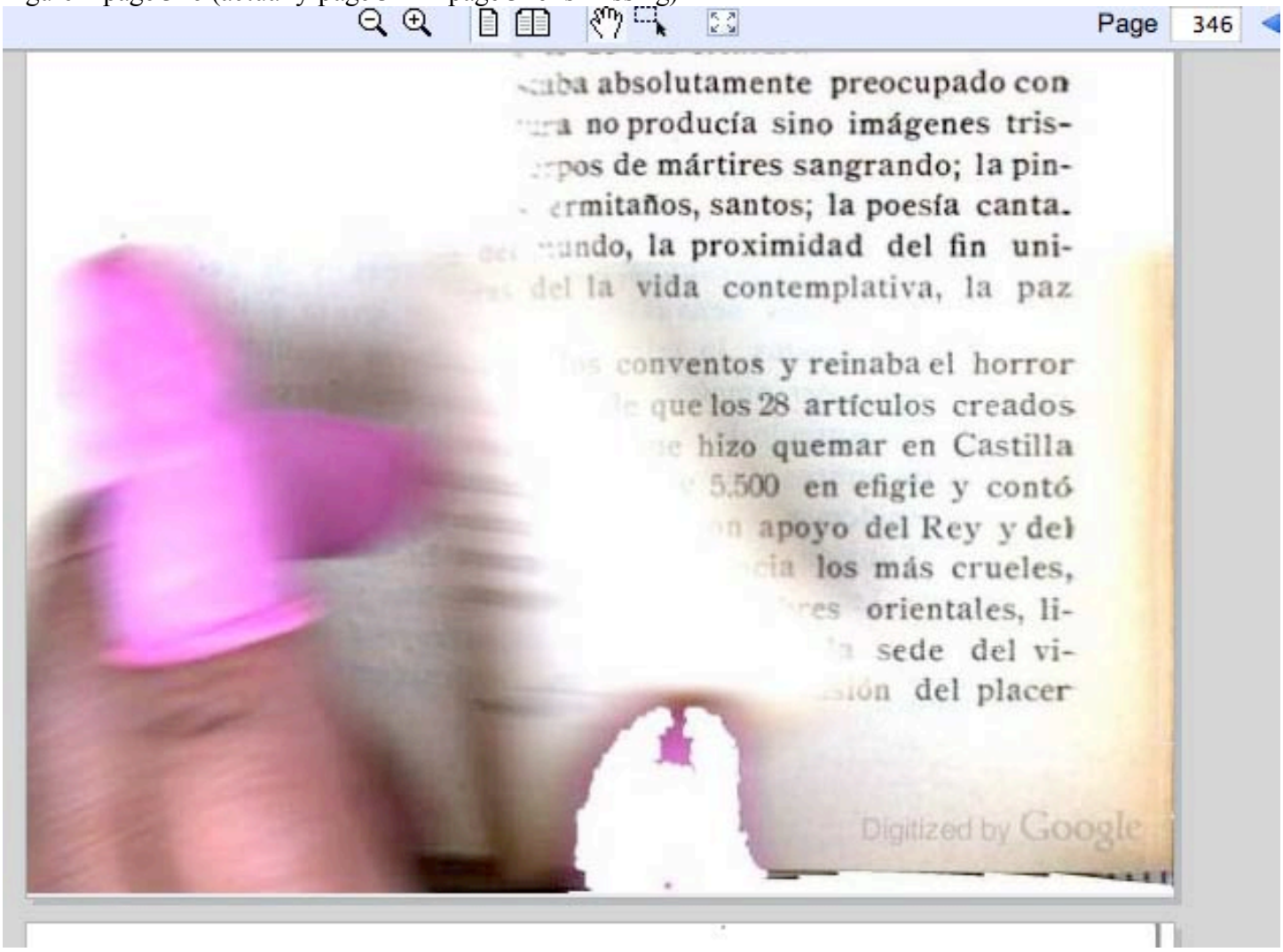


Figure 5 page 347 (the second page 347)

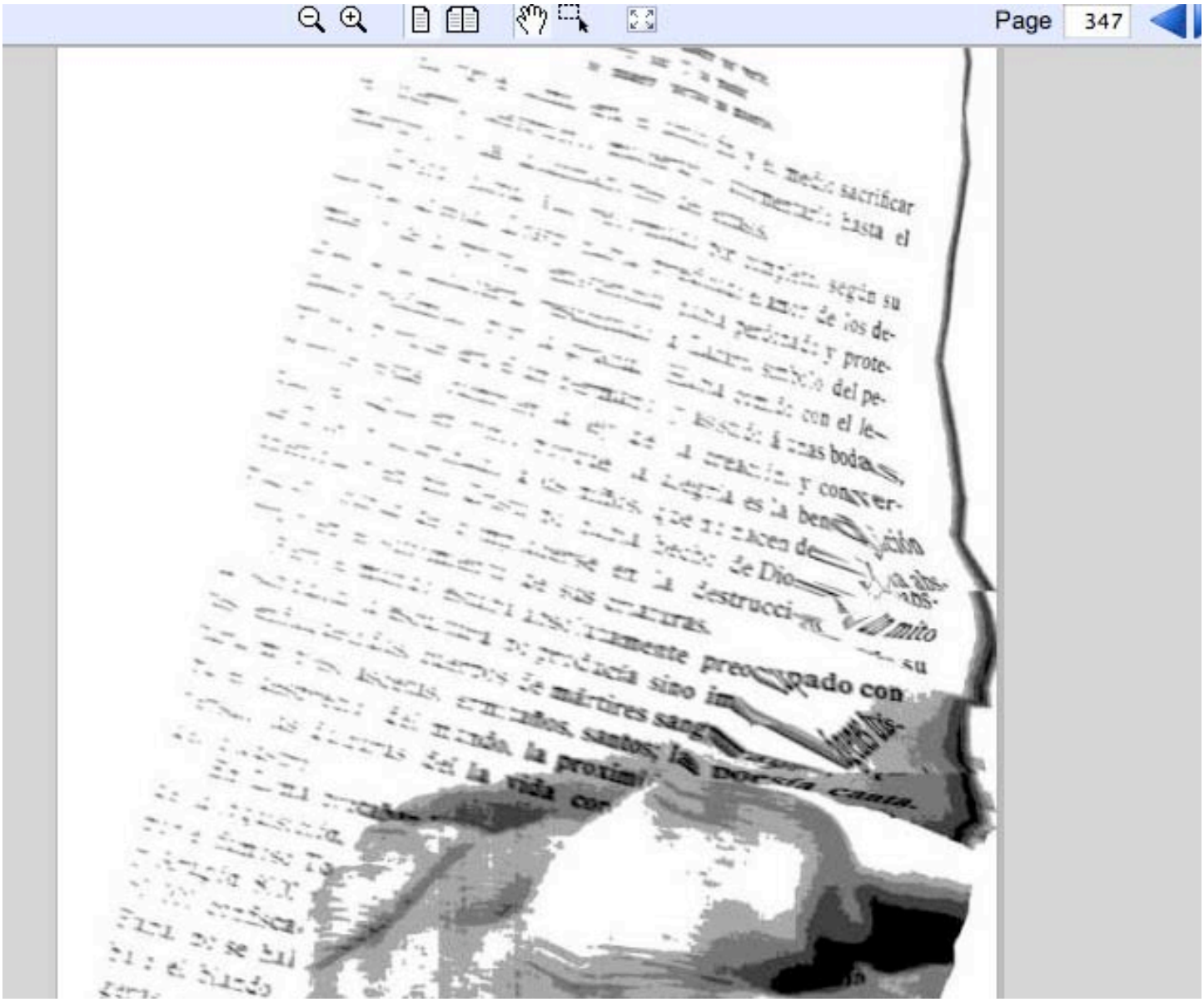
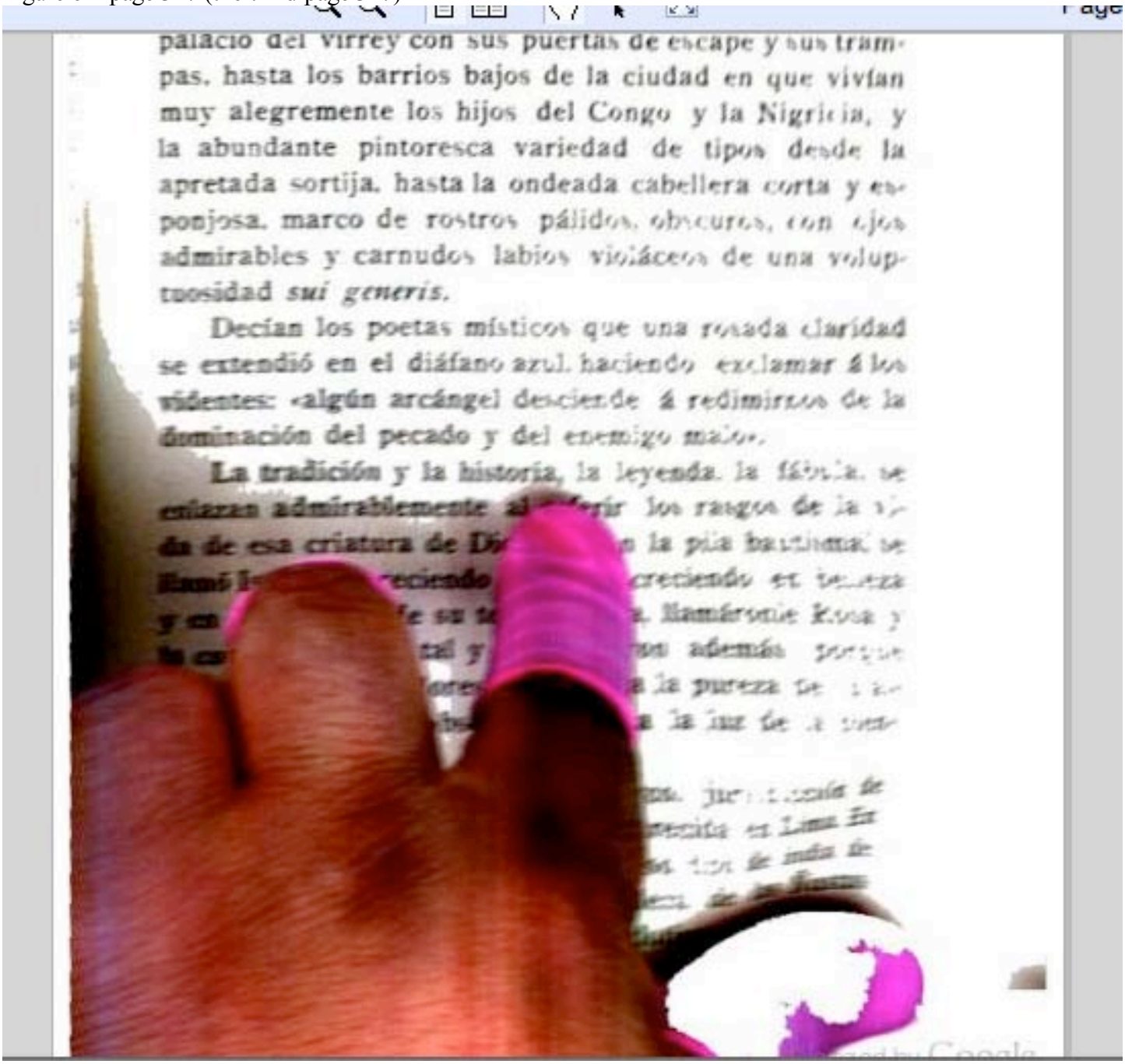


Figure 6 – page 347 (the third page 347)



palacio del virrey con sus puertas de escape y sus trampas, hasta los barrios bajos de la ciudad en que vivían muy alegremente los hijos del Congo y la Nigricia, y la abundante pintoresca variedad de tipos desde la apretada sortija, hasta la ondeada cabellera corta y esponjosa, marco de rostros pálidos, obscuros, con ojos admirables y carnudos labios violáceos de una voluptuosidad *sui generis*.

Decían los poetas místicos que una rosada claridad se extendió en el diáfano azul, haciendo exclamar á los videntes: «algún arcángel desciende á redimirnos de la dominación del pecado y del enemigo malo».

La tradición y la historia, la leyenda, la fábula, se enlazan admirablemente al referir los rasgos de la vida de esa criatura de Di[illegible] la pila bautismal se llamó [illegible] creciendo [illegible] creciendo en belleza y en [illegible] su [illegible] llamáronle Rosa y la [illegible] tal y [illegible] además porque [illegible] la pureza de [illegible] la luz de la [illegible]

[illegible] jurisdicción de [illegible] en Lima [illegible] de [illegible]

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt **Date of investigation**: August 10, 2008

Book title: Lof Lýginnar
Author: Porleifur Halldórsson
Year of Publication: 1915
Place of Publication: Cornell University Library
ISBN number?: na
Content/Disciplinary area: Icelandic, Icelandic literature and language

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	Provided by Stanford University, seemingly good quality
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Metadata given by Google minimal- book sections are given. This book is three books in one, with the pagination starting over after each book. Confusion can take place if, for example, a scholar enters page '10' – only page 10 of the first book comes up. It is not clear whether these are three volumes or separate texts?
MARC record available?	na
Can the book be searched through an OPAC? If so, which one?	no
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	Accurate but not a lot of information

SCAN SEARCH

search paradigm settings (example: every 10 double-pages): scanned through entire book

Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
	Excellent quality- no anomalies in this book. Scan is very professionally done			

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	**Comment**
	(Is OCR search permitted): Y / N
prýðelegustu	unsuccessful detection – even though this word appears on page 7

ofannkomner	successful detection
giörer	successful detection
Fishcheriæ	unsuccessful detection (page 53)
Breiðabólsstaður	unsuccessful detection
formaala	unsuccessful detection
ætiid	unsuccessful detection
samsettr	successfully detected
vtlagde	successfully detected
bókahnútur	successfully detected

IMAGE ASSESSMENT

If the original text includes graphics/illustrations, are they included in the electronic version?	Yes, on page 2, 6, 10, 14, 17, 28, 29, 31, 35, 40, 58, 61, 65 in second book; page 43 in third book. See Figure 1 for an example.
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	No
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	Not sure- I think it's a PDF. No writing within any of the texts is OCR recognizable.
Is the scan on a gray-scale or color scale?	Grey
Quality – does it look like the real thing?	Yes
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	Yes, absolutely
Are the illustrations searchable?	No
Ability to print illustrations? (is there a watermark?)	Yes, no watermark
Are the images downloadable? (file size, pixel quality)	Yes, if you download the entire document
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	zooming yes, but this could be done with a microscope and an analog version. Other than that, there are no differences between digital and analog for the images in this text.
Other notes	

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

GENERAL OBSERVATIONS

Comment on ease of discovery in integrated interface (OPAC or better)	Easily accessible in Google Books
Overall quality of scan of this book	EXCELLENT
Ease of access to full text	Yes
Ability to copy, quote, cite?	Yes
Ability to search?	Yes
Ability to download?	Yes
Ability to print?	Yes
Scanned book allow for consulting of contents?	Yes, but some Icelandic characters are not recognizable. This is a limitation for research with the Icelandic language.
How successful is OCR recognition?	I´ve determined that OCR recognition is successful for all Icelandic characters except for æ, ð, Æ, and Ð. Otherwise, it accepts accent marks and umlauts.
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Absolutely
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Definitely
Issues with this book that would affect usefulness for scholars	Some Icelandic characters cannot be detected by OCR. Also, because this text is three books in one, the pagination system – and it's searchability – are problematic. For example, the pagination of Book 2 starts over after Book 1 – same goes for Book 3. Therefore, there are THREE page 10s, page 11s, page ix, etc. If you enter the page number to jump to that page, it will only work for Book 1. Therefore, to get to duplicate page numbers of either Book 2 or 3, the scholar must scroll through the entire text to find them.
Can this book be shared with other scholars/students?	Yes, by emailing the URL
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	NA
Other notes	

Book URL: http://books.google.com/books?id=3AIUAAAAIAAJ&pg=PR3&dq=grammar+history&lr=lang_is&as_brr=1#PPP2,M1

Date accessed: Aug 10, 2008

Figure 1 – example of image from this book

Coat-of-arms of Iceland, from the *Psalma Bok*, 1589.

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt **Date of investigation**: Aug 10, 2008

Book title: Subsidios para a historia de Macau
Author: Bento da França
Year of Publication: 1888
Place of Publication: Imprensa Nacional, Portugal
ISBN number?: NA
Content/Disciplinary area: Macau, history, Mandarin and Portuguese

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	Seemingly good condition. provided by Stanford University
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Easily accessibly on Google Books. Some metadata is provided by Google Books, and it includes book sections, key words, and other editions by this same author.
MARC record available?	NA
Can the book be searched through an OPAC? If so, which one?	No
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	Pretty accurate

SCAN SEARCH

search paradigm settings (example: every 10 double-pages) Scrolled through entire book

Page(s):	readability/ quality of scan	anomalies?	other?	other?
1, 6, 8, 18, 20, 22, 30, 36, 50, 56, 60, 70, 74, 84, 92, 98, 102, 118, 124, 130, 132, 138, 142, 148, 162, 164, 174, 184,		Scanners fingertips completely visible – see Figure 1 as an example	The ubiquitous nature of this can be slightly distracting to the eye	
26	page was moved during scan, illegible (Figure 2)			
30	very light, page was slightly moved during scan. Some words illegible			
36, 44	pages were moved during scan – completely illegible			
53	ILLEGIBLE.	This page is present two times in this digital book- both times with the previous page being	See Figure 3	

		turned while scan took place.
68, 72, 90, 92, 93, 94, 98, 140, 146, 148, 196, 210	parts (or all) of these pages are completely illegible	94 and 210 are quite extreme

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	Comment (Is OCR search permitted): Y / N
Macau	successful recognition
portugués	no hits
china	successful recognition
Marco Polo	successful recognition
lingua	successful recognition
gente	successful recognition
mandarins	successful recognition
imperador	successful recognition
jesuita	successful recognition
não, tão	successful recognition

IMAGE ASSESSMENT

If the original text includes graphics/illustrations, are they included in the electronic version?	No images in this book
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	
Is the scan on a gray-scale or color scale?	
Quality – does it look like the real thing?	
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file	

size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	
Other notes	

GENERAL OBSERVATIONS

Comment on ease of discovery in integrated interface (OPAC or better)	Easily accessible in Google Books
Overall quality of scan of this book	Very poor
Ease of access to full text	Yes
Ability to copy, quote, cite?	Yes
Ability to search?	Yes
Ability to download?	Yes
Ability to print?	Yes
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	seems to be pretty accurate, except for those pages whose words are so blurred that scanning is impossible
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	No – would recommend complete re-scan of this book due to poor quality
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	No
Issues with this book that would affect usefulness for scholars	Significant amount of pages that are illegible
Can this book be shared with other scholars/students?	Yes
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	NA
Other notes	

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Date accessed: August 9, 2008

Figure 1 Scanner Fingers appearing in book scan

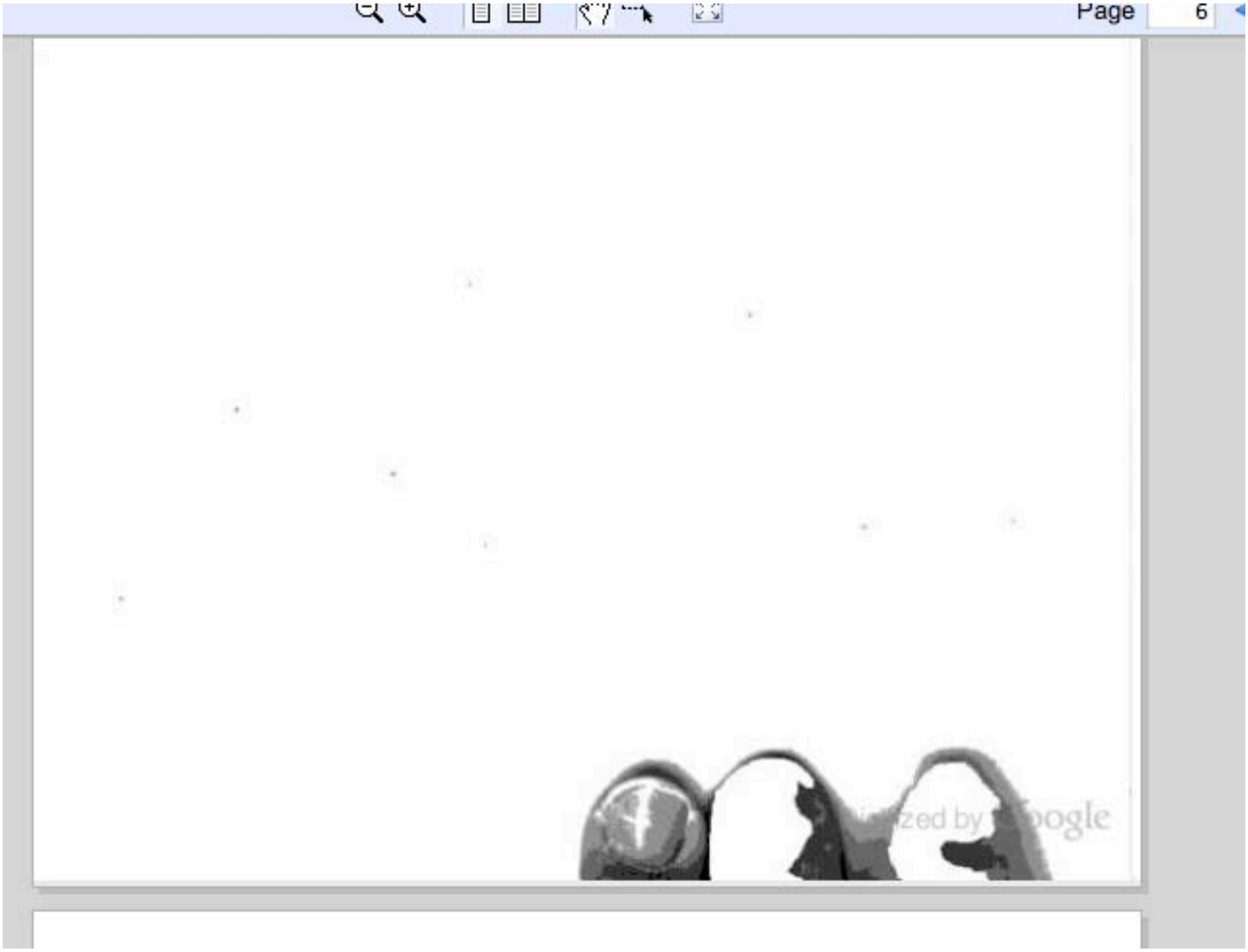
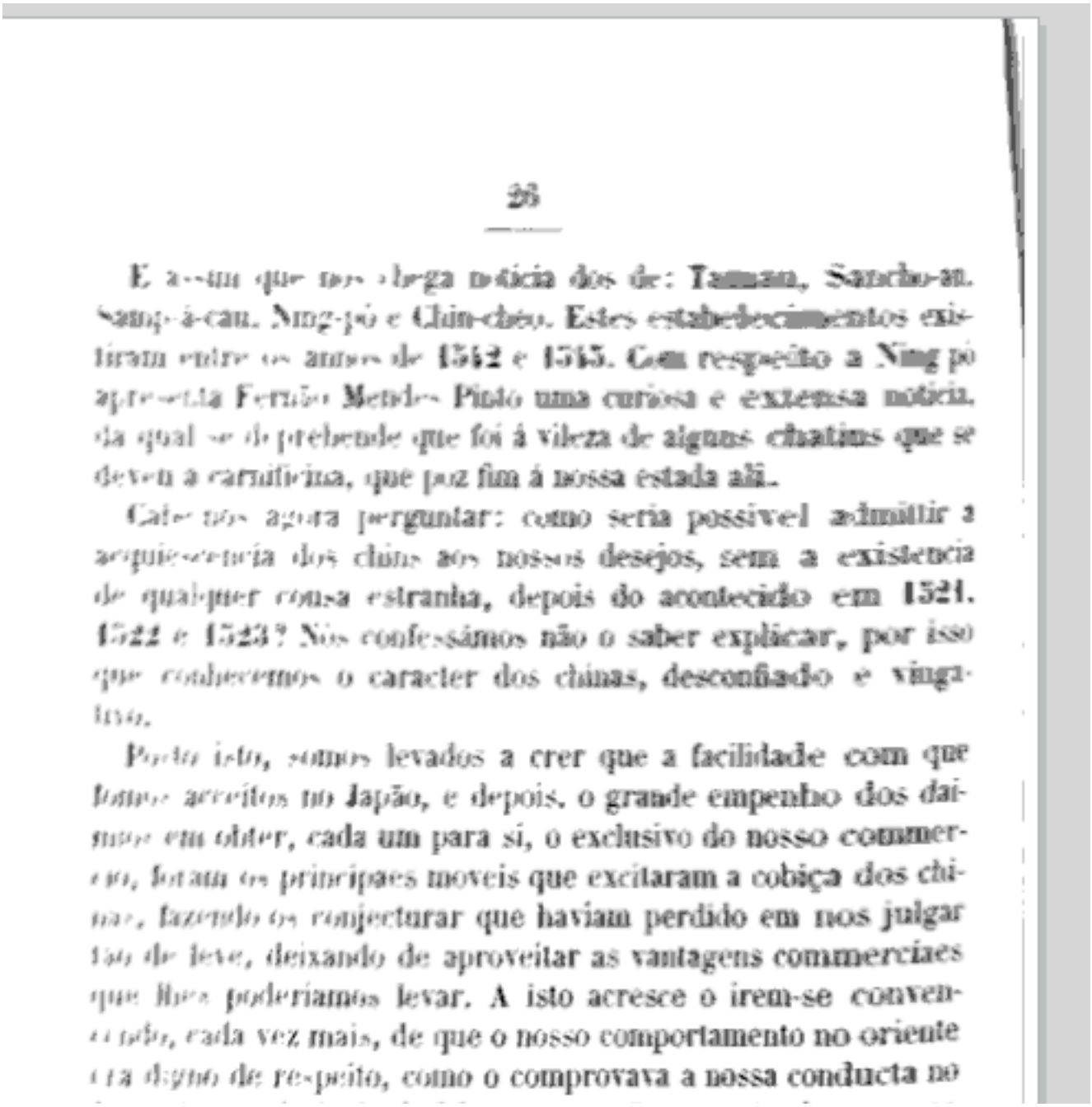


Figure 2 Page 26 – moved during scan



26

E assim que nos chega noticia dos de: Tamau, Sancho-an, Samp-á-cau, Ning-pó e Chin-cheo. Estes estabelecimentos existiram entre os annos de 1542 e 1545. Com respeito a Ning pó apresenta Fernão Mendes Pinto uma curiosa e extensa noticia, da qual se deprehende que foi á vileza de alguns chatins que se deveu a carnificina, que poz fim á nossa estada ali.

Cabe-nos agora perguntar: como seria possivel admittir a acquiescencia dos chins aos nossos desejos, sem a existencia de qualquer cousa estranha, depois do acontecido em 1521, 1522 e 1523? Nós confessámos não o saber explicar, por isso que conhecemos o caracter dos chinas, desconfiado e vingativo.

Posto isto, somos levados a crer que a facilidade com que fomos acceitos no Japão, e depois, o grande empenho dos daimios em obter, cada um para si, o exclusivo do nosso commercio, foram os principaes moveis que excitaram a cobiça dos chinas, fazendo-os conjecturar que haviam perdido em nos julgar tão de leve, deixando de aproveitar as vantagens commerciaes que lhes poderiamos levar. A isto acresce o irem-se convencendo, cada vez mais, de que o nosso comportamento no oriente era digno de respeito, como o comprovava a nossa conducta no

Figure 3 Page 53 (first page 53)

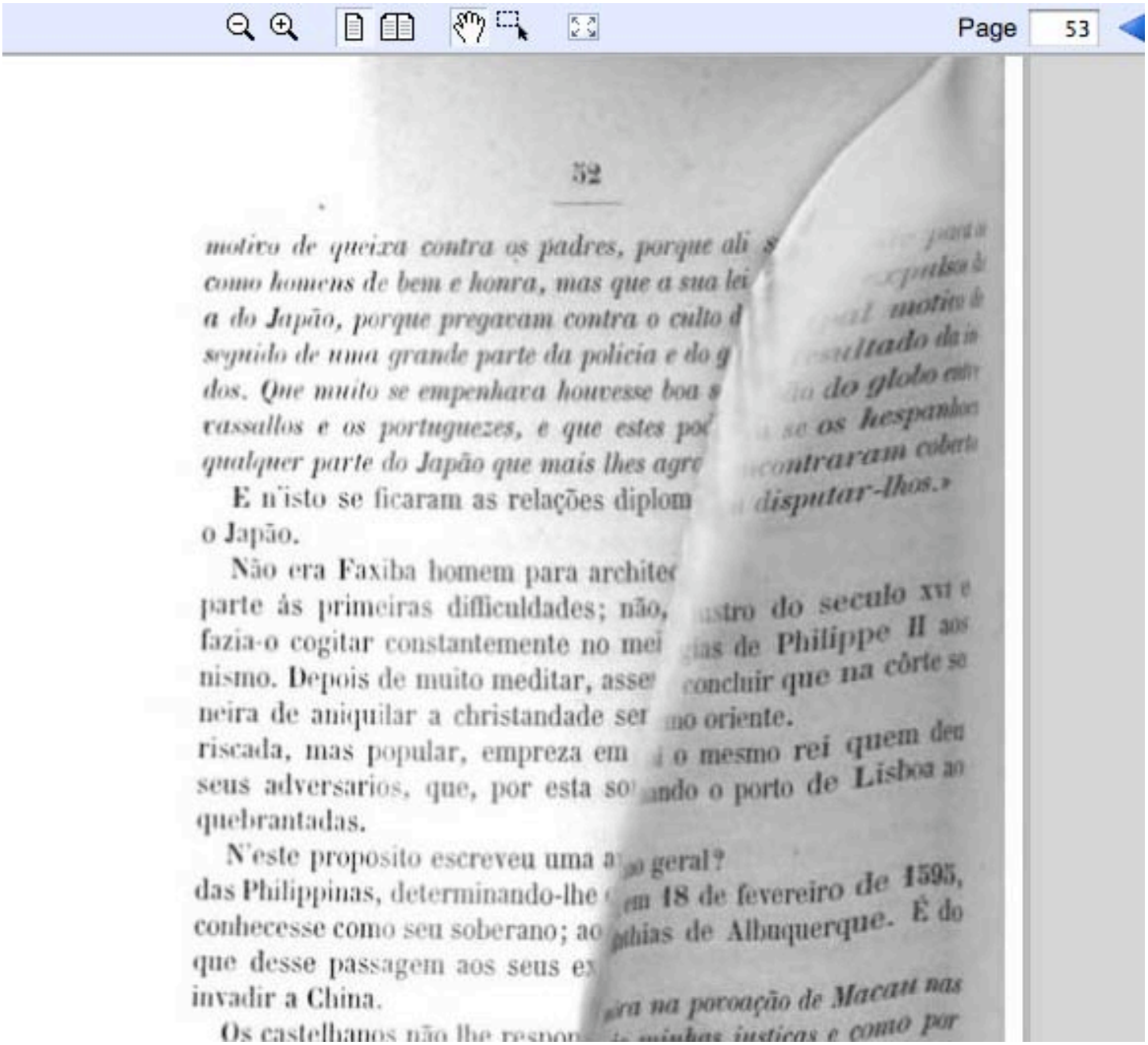
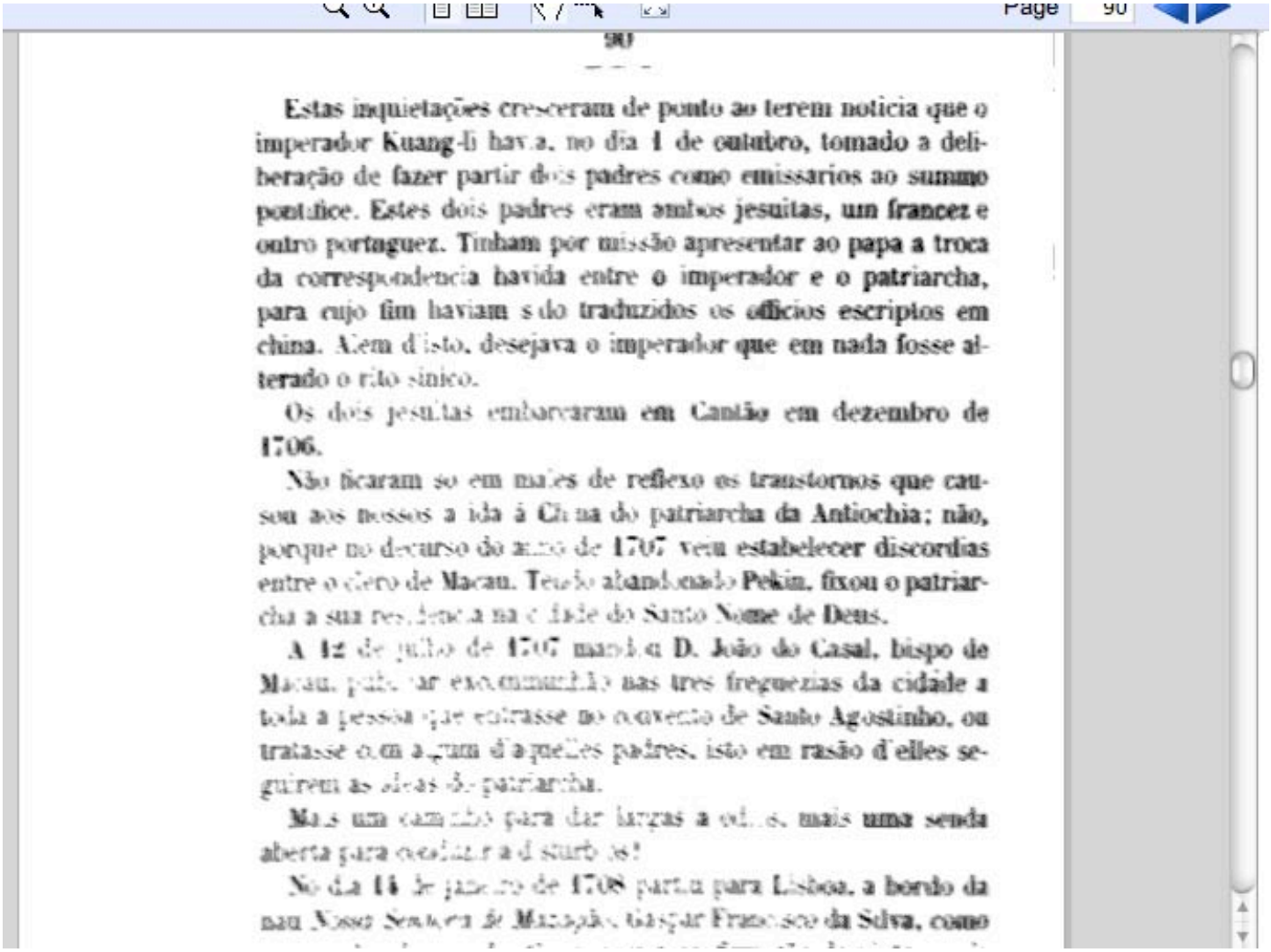


Figure 4 Example of quality of scanning in this book



90

Estas inquietações cresceram de ponto ao terem noticia que o imperador Kuang-li havia, no dia 1 de outubro, tomado a deliberação de fazer partir dois padres como emissarios ao summo pontifice. Estes dois padres eram ambos jesuitas, um francez e outro portuguez. Tinham por missão apresentar ao papa a troca da correspondencia havida entre o imperador e o patriarcha, para cujo fim haviam sido traduzidos os officios escriptos em china. Alem d'isto, desejava o imperador que em nada fosse alterado o rito sinico.

Os dois jesuitas embarcaram em Cantão em dezembro de 1706.

Não ficaram só em males de reflexo os transtornos que causou aos nossos a ida á China do patriarcha da Antiochia; não, porque no decurso do anno de 1707 veiu estabelecer discordias entre o clero de Macau. Tendo abandonado Pekin, fixou o patriarcha a sua residencia na cidade do Santo Nome de Deus.

A 12 de julho de 1707 mandou D. João do Casal, bispo de Macau, publicar excommunhão nas tres freguezias da cidade a toda a pessoa que entrasse no convento de Santo Agostinho, ou tratasse com algum d'aquelles padres, isto em rasão d'elles seguirem as ideas do patriarcha.

Mais um caminho para dar largas a odios, mais uma senda aberta para conduzir a disturbios!

No dia 14 de janeiro de 1708 partiu para Lisboa, a bordo da nau Nossa Senhora de Mazagão, Gaspar Francisco da Silva, como

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt **Date of investigation**: August 17, 2008

Book title: Parnaso Cubano
Author: Antonio López Prieto
Year of Publication: 1881
Place of Publication: Habana, Cuba. Miguel de Villa, Librería
ISBN number?:
Content/Disciplinary area: Cuban literature, Cuban poetry

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	Original from University of Wisonsin-Madison
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Easily accessible from Google Books. Metadata provided by Google sufficient.
MARC record available?	NA
Can the book be searched through an OPAC? If so, which one?	No
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	NA – no table of contents

SCAN SEARCH

search paradigm settings (example: every 10 double-pages) scrolled through entire book

Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
xlviii, lxiv	text on this page is so light that parts of it are hard to read. In fact, word searches of this part do not yielded any search results.			
9-10	extreme lightness of text on these pages, hard to read, words not detected in searches			
31	this page is so light it is illegible on left side	(see figure 1)		
no other major anomalies, besides lightness of scan				

WORD SEARCHES DONE WITH THIS BOOK

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	Comment (Is OCR search permitted): Y / N
UDELIQUIA	successful recognition
desierto	successful recognition
siniestra	successful recognition
metatésis	unsuccessful – this word appears in the footnote on page 5, but what not detected
Daoiz	successful recognition
la madre Vesta	successful recognition
Zequeira	successful recognition
señor	successful recognition
trágica	successful recognition
Retrátanse	unsuccessful – this word appears in the footnote on page 65, but what not detected

IMAGE ASSESSMENT

If the original text includes graphics/illustrations, are they included in the electronic version?	No images in this text
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	
Is the scan on a gray-scale or color scale?	
Quality – does it look like the real thing?	
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	
Other notes	

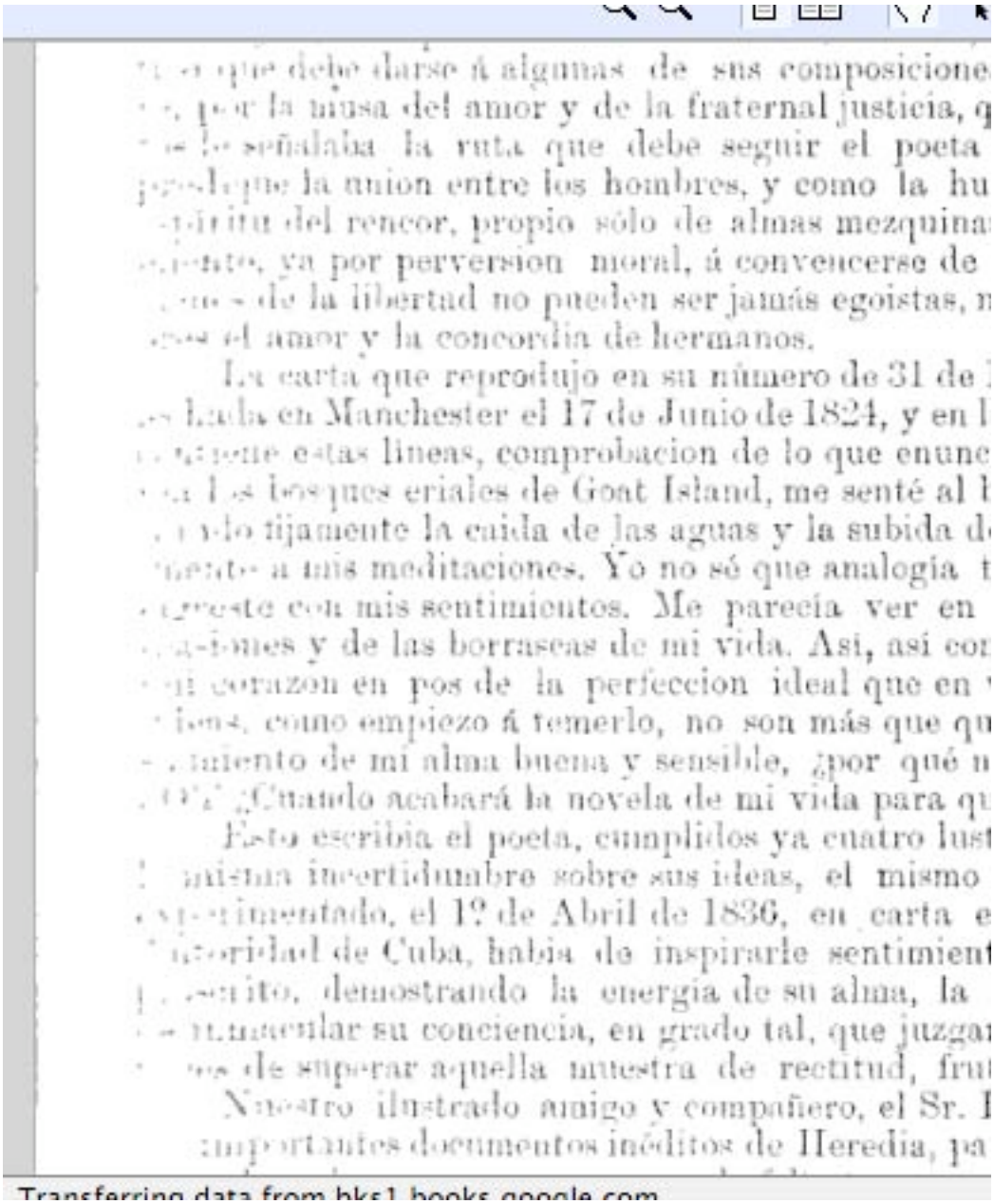
GENERAL OBSERVATIONS

Comment on ease of discovery in	Easily accessible in Google books

integrated interface (OPAC or better)	
Overall quality of scan of this book	Good, but some major lightness of scan. This sometimes makes the text illegible.
Ease of access to full text	Yes
Ability to copy, quote, cite?	Yes
Ability to search?	Yes
Ability to download?	Yes
Ability to print?	Yes
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	sometimes good, sometimes not. This may be due to the fact that the scan quality (extreme lightness) affects readabily of words sometimes.
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Possibly
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Possibly
Issues with this book that would affect usefulness for scholars	Some words are not detectable in this book
Can this book be shared with other scholars/students?	Yes, by emailing the URL
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	No items missing
Other notes	

Book URL:
http://books.google.com/books?id=nrNGAAAAMAAJ&pg=PA146&dq=adquisicion+de+lengua+por+ninos&lr=&as_brr=1#PPA31,M1
Date accessed: August 17, 2008

Figure 1, example of extreme lightness present often in this text



CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt **Date of investigation**: August 16, 2008

Today I wanted to explore some books written by Jean Piaget, a historic and well-known scholar on child L1 development and psychology, whose research has led a significant amount of research on child language acquisition, both first and second languages. I was at first disappointed at how much of Piaget's material is available on Google books, but is not available for preview. An advanced search setting, designated at only yielded books for full view, do not provide any of his texts. I'll therefore analyze one of his texts available for limited preview – I imagine it must be a copyright issue.

Ok, after some searching, I see that none of his books, particularly Language and Thought and on children's judgment, are available for even partial view- only snippet. I'll do a new edition/translation of one of his texts, The Child's Conception of Physical Causality, done in 1929.

Book title: The Child's Conception of Physical Causality
Author: Jean Piaget, Jaan Valsiner (introduction by Valsiner)
Year of Publication: original 1929, this edition published in 2001
Place of Publication: Transaction Publishers
ISBN number?: 076580641X
Content/Disciplinary area: child psychology

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	Seemingly good quality, this book is limited preview only, however
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Limited preview on Google; unfortunately not any of Piaget's originals are available for full view
MARC record available?	NA
Can the book be searched through an OPAC? If so, which one?	Yes! Georgetown University, as an example, makes this book available electronically in their ebrary. See URL at http://0-site.ebrary.com.library.lausys.georgetown.edu/lib/georgetown/Doc?id=10017131
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	Full metadata available, content, keywords, etc. Detailed.

SCAN SEARCH

search paradigm settings (example: every 10 double-pages)	Searched through entire book (or those pages rather that were available for preview)

Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
Xiii, xxi, 5, 7, 11, 13,	Section title and/or top of page wording is blurred. However, it is only one line of text and thus does not prevent readabilty			
Every page after 16 is not available for				

view

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	Comment (Is OCR search permitted): Y / N
language	Successful recognition
awareness	Successful recognition, however in pages not accessible for my view
peer	Successful recognition
reasoning	Successful recognition, however in pages not accessible for my view
Development	Successful recognition, however in pages not accessible for my view
Adolescent	Successful recognition, however in pages not accessible for my view
Output	No hits
Acquisition	No hits
Conversation	Successful recognition
Speech	Successful recognition, however in pages not accessible for my view

IMAGE ASSESSMENT

If the original text includes graphics/illustrations, are they included in the electronic version?	No images in this text
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	
Is the scan on a gray-scale or color scale?	
Quality – does it look like the real thing?	
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	
Other notes	

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

GENERAL OBSERVATIONS

Comment on ease of discovery in integrated interface (OPAC or better)	Limited preview in Google Books. Very few pages are available for view, of those 16 available, the top line of several of the pages is blurred.
Overall quality of scan of this book	Ok
Ease of access to full text	Yes
Ability to copy, quote, cite?	Yes
Ability to search?	Yes, though limited preview
Ability to download?	No
Ability to print?	No
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	Seemingly perfect
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Yes, in full view status perhaps
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes, in full view status
Issues with this book that would affect usefulness for scholars	Not available for preview
Can this book be shared with other scholars/students?	Yes, by emailing the URL
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	NA
Other notes	Vii and viii, the Table of Contents, have wonderful hyperlinks for researchers to jump to specified sections in book

Book URL: http://books.google.com/books?id=_RUQHiX-5HoC&printsec=frontcover&dq=jean+piaget&lr=&sig=ACfU3U0iOqIQ7dQoHXNyyN-vASp6bCGbiw#PPA7,M1

Date accessed: August 16, 2008

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt **Date of investigation**: August 16, 2008

Book title: Child's Conception of Physical Causality
Author: Jean Piaget
Year of Publication: 1999
Place of Publication: Routledge (reprinted; London). Original was published in 1930 by Kegan Paul, Trench, Trubner & Co. Ltd.
ISBN number?: 0415-20998-6
Content/Disciplinary area: child development, child psychology

AVAILABILITY SEARCH

Source of scanned book:	Georgetown University Ebrary
Comments on condition/provenance of book:	Seemingly excellent quality, available for full access view from Georgetown's Ebrary.
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Easily accessible by a link in the GEORGE system. Excellent hyperlinked sections available in Table of Contents. Scrolling capability also nice.
MARC record available?	NA
Can the book be searched through an OPAC? If so, which one?	Yes – Georgetown's Ebrary. Metadata provided is excellent.
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	Table of contents is all hyperlinked- fantastic!

SCAN SEARCH

search paradigm settings (example: every 10 double-pages)	Searched through entire book			
Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
	Absolutely NO anomalies in this book, outstanding quality. This book is perfect example of how a digitized text should be.			

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	Comment (Is OCR search permitted): Y / N
reasoning	Successful detection. Highlights findings in red
Awareness	Successful detection
Causal reasoning	No hits
Egocentric logic	Successful detection
language	Successful detection, however this word was in the footnote and was not highlighted well by highlighter tool.

acquisition	No hits
consciousness	Successful detection
Interaction	Successful detection, allows me to scroll through only those pages with this term – wonderful search tool
adult	Successful detection
peer	No hits

IMAGE ASSESSMENT

If the original text includes graphics/illustrations, are they included in the electronic version?	No images, but on pages 199 and 200 there are hand drawn images done by children to depict how a bicycle works. Neat to see this.
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	
Is the scan on a gray-scale or color scale?	
Quality – does it look like the real thing?	
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	
Other notes	

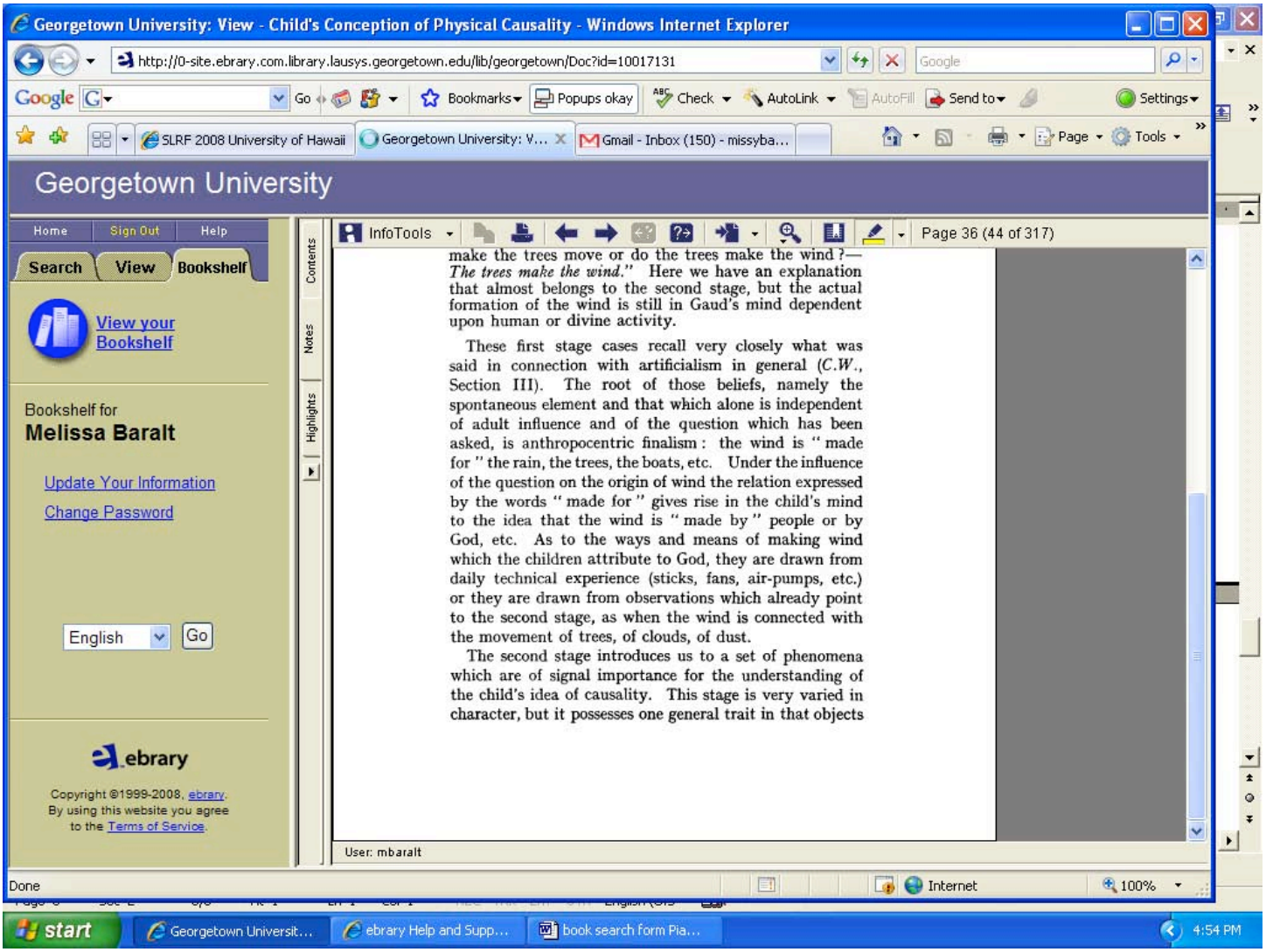
GENERAL OBSERVATIONS

Comment on ease of discovery in integrated interface (OPAC or better)	This book is available for full access through Georgeown's Ebrary. It was presented as a link when I searched in Georgetown's OPAC system for the text. I had to create an account to be able to use some of the tools in Ebrary (i.e. highlighter), but otherwise could see the entire book without logging in.
Overall quality of scan of this book	FANTASTIC – this is an example of what a digitized book should look like. No

	anomalies whatsoever, very professional. No colorations, scanning issues as Google typically has.
Ease of access to full text	100% accessible
Ability to copy, quote, cite?	Yes, very unique tools for research that aren't available in Google. For example, I can use a highlighter tool, choose highlighter color, take notes digitally and keep them in my account with the book, tag pages, and copy text and paste it into my notes if I want. I can also copy text and transfer and paste to a word doc. This is not permissible in Google, for example. The Ebrary interface is also very user friendly.
Ability to search?	Yes
Ability to download?	No, but I can keep my own digital books on my digital 'bookshelf' for any future reference.
Ability to print?	Yes
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	100% perfect
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Absolutely
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes
Issues with this book that would affect usefulness for scholars	Ideal platform for researchers, as the Ebrary interface provides unique tools pertinent for any scholar – note taking, highlighting in different colors, tagging of text and pages, and saving notes and comments.
Can this book be shared with other scholars/students?	Yes
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	NA
Other notes	Fantastic digital book system

Book URL: http://0-site.ebrary.com.library.lausys.georgetown.edu/lib/georgetown/Doc?id=10017131
(not sure if this will work, one might have to log into Georgeotown's library system for accesibilty)
Date accessed: August 16, 2008

Example of Ebrary interface (I did this one on a PC)



CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt	**Date of investigation**: August 22, 2008

Book title: Portugual; diccionario historico, chorographico, biographico, bibliographico, heraldico, numismatico e artistico

NOTE: Google Books lists the title as "Portugual; diccionario historico, chorographico, **heraldico**, biographico, bibliographico, numismatico e artistico", which is an error of ordering of the words.

Author: João Manuel Esteves Pereira, Guilherme Rodrigues
Year of Publication: 1907
Place of Publication: João Romano Torres & Co.
ISBN number?: NA
Content/Disciplinary area: history of Portugal

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	original provided by University of Michigan
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Easily accessible in Google Books, metadata provided includes content and pictures
MARC record available?	NA
Can the book be searched through an OPAC? If so, which one?	No
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	Metadata provided is sufficient

SCAN SEARCH

search paradigm settings (example: every 10 double-pages)	every 10 double pages

Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
6, 245, 273, 283, 498, 500, 542	scan cuts off text on far side of page			
714	text in footnote is so small and blurry that there is no OCR recognition of any of the words			

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	Comment
	(Is OCR search permitted): Y / N
Lisboa	successful recognition
Dardavaz	successful recognition
Ericeira	successful recognition
Manuel Affonso	successful recognition
Sermão	successful recognition
devoção	successful recognition
Parada de Gonta	successful detection on page 415, but not on page 300 – the words are so blurred on 300 that it was not detected.
Rio Lima	successful recognition
Moçambique	successful recognition
ascensões	successful recognition

IMAGE ASSESSMENT

If the original text includes graphics/illustrations, are they included in the electronic version?	Yes – many are drawings and photos (see Figure 1 for example)
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	No
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	OCR – good
Is the scan on a gray-scale or color scale?	gray-scale
Quality – does it look like the real thing?	yes – most of the images are of high quality, some are poorer. This may be due to the age of the original book
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	Yes
Are the illustrations searchable?	only by the picture titles
Ability to print illustrations? (is there a watermark?)	No watermark – pictures are printable
Are the images downloadable? (file size, pixel quality)	Only by downloading the entire document.
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	No
Other notes	

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

GENERAL OBSERVATIONS	
Comment on ease of discovery in integrated interface (OPAC or better)	Easily accessible in Google books
Overall quality of scan of this book	Average - many of the pages have text that is cut off by the scanner – especially text on far side of page
Ease of access to full text	Yes
Ability to copy, quote, cite?	Yes
Ability to search?	Yes
Ability to download?	Yes
Ability to print?	Yes
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	Good except for smaller and blurred text. This is due to scan or possible original condition of book
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Yes
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes
Issues with this book that would affect usefulness for scholars	Some searching issues, however overall a good references
Can this book be shared with other scholars/students?	Yes, by emailing the URL
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	NA
Other notes	This book has some very interesting images and pictures

Book URL:
http://books.google.com/books?id=R5cfAAAAMAAJ&dq=historia+de+mo%C3%A7ambique&lr=lang_pt&as_brr=1&source=gbs_summary_s&cad=0

Date accessed: August 22, 2008

Figure 1 example of image in text

estrangeiro. N'esta freg., junto ao rio Lima, estão as ruinas d'uns antiquissimos paços dos duques de Bragança. Darque pertence á 3.ª div. mil. e ao distr. de recrut. e res. n.º 3, com a séde em Vianna do Castello.

Dasbaparà. Pov. da praganã de Nagar-Avely, no distr. de Damão, prov. e arceb. de Gôa, India.

Dasiga. Terras do distr. de Sofala, na prov. de Moçambique, Africa Oriental, limitrophes das terras Pongue e muito aptas para toda a qualidade de cultura. Tem muitas palmeiras e matto. N'estas terras estão edificadas varias povoações.

Daugim. Baluarte da antiga cidade de Gôa, India, edificado em 1565 pelo 9.º vice-rei, D. Antão de Noronha, e abandonado em principios do seculo XVIII pelo 48.º vice-rei e capitão-general Bernardo José de Lorena, conde de Sarzedas. Fazia parte do systema de fortificações comprehendidas na grande muralha da margem oriental da ilha de Gôa para defeza d'esta e da cidade de Gôa.

Daugravary-grande. Pov. da propriedade rustica do mesmo nome, na freg. de Santo André de Brancavará, do conc. de Diu, na India.

Daugravary-pequeno. Pov. da propriedade rustica do mesmo nome, na freg. de Santo André de Brancavará, no conc. de Diu, na India.

Daun e Lorena *(Manuel Joaquim Saldanha d'Oliveira).* Capitão do corpo d'officiaes da administração militar, ao serviço do ministerio da marinha, e actualmente secretario do governo do districto de Mossamedes. N. em 29 de novembro de 1859, assentou praça em 16 de maio de 1881, foi promovido a alferes em 2 de janeiro de 1885, a tenente em 7 de fevereiro de 1895, e a capitão em 14 de junho do mesmo anno. E' descendente do fallecido marechal duque de Saldanha. Tomou parte nas operações de guerra em Moçambique, realisadas em 1894 e 1895, como chefe da delegação da administração militar em Lourenço Marques. Embarcou para a provincia de Angola em 6 de junho de 1899, exercendo até 24 de agosto

1.º Conde de Daupias

gou á maioridade. Tornando-se, por herança de seus paes, proprietario da fabrica de lanificios ao Calvario, em Alcantara, conseguiu pela sua actividade e elevada competencia na direcção administrativa, que aquella industria se desenvolvesse e se engrandecesse, tornando-se a fabrica a primeira do seu genero que existiu em Portugal. Os avultados lucros, que auferia, fôram uma fonte de riqueza para o seu proprietario. Além dos contramestres e empregados dos escriptorios, aquellas officinas sustentavam centenares de operarios, que eram largamente compensados, e que o conde Daupias tratava com estima e conside-

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt **Date of investigation**: August 16, 2008

After continued searches I finally found Piaget's Psychology of the Child, however for limited preview only. This version is a reprint with new translation and forward.

Book title: Psychology of the Child
Author: Jean Piaget
Year of Publication: 1972
Place of Publication: Basic Books
ISBN number?: 0465095003
Content/Disciplinary area: child psychology and development

AVAILABILITY SEARCH

Source of scanned book:	Good Books
Comments on condition/provenance of book:	Seemingly good condition, except for the fact that it is limited preview. Therefore, any scholar does not have full access to all of the pages due to copyright issues.
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	This is not an immediate search result; I had to look through several pages of Google search yields to find it. It also helps to do a search for full previews only. Metadata provided by Google is sufficient.
MARC record available?	NA
Can the book be searched through an OPAC? If so, which one?	No, however, Georgetown's GEORGE system does provide an insider view where the table of contents is shown
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	Sufficient. The table of contents in this book is completely hyperlinked, which is a wonderful resource for scholars. This makes it easy to jump to these specific sections in the book.

SCAN SEARCH

search paradigm settings (example: every 10 double-pages) searched through every page in the book

Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
2, all pages after 28	These pages are not available for view			
23, 25,	blurriness in section title, does not impede readability however			

WORD SEARCHES DONE WITH THIS BOOK

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	Comment (Is OCR search permitted): Y / N
language acquisition	successful detection, but not available for preview
cognitive aspect, affective aspect	successful detection, but not available for preview
symbolic function	successful detection
reason	successful detection
awareness	successful detection
input	successful detection
assimilation	successful detection
sensori-motor intelligence	successful detection
speech	successful detection
stage	successful detection

IMAGE ASSESSMENT

If the original text includes graphics/illustrations, are they included in the electronic version?	No images in this book
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	
Is the scan on a gray-scale or color scale?	
Quality – does it look like the real thing?	
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	
Other notes	

GENERAL OBSERVATIONS

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Comment on ease of discovery in integrated interface (OPAC or better)	Not easily accessible in Google Books, this book is unfortunately limited preview only
Overall quality of scan of this book	The pages available for view have excellent quality scans
Ease of access to full text	No
Ability to copy, quote, cite?	No
Ability to search?	Yes
Ability to download?	No
Ability to print?	No
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	so far, perfect search and detection capabilities
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	No as not whole book is not included
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	No, whole book is not available
Issues with this book that would affect usefulness for scholars	Limited preview
Can this book be shared with other scholars/students?	Yes, by emailing the URL
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	NA
Other notes	Though limited preview, the table of contents are provided, and searching of the entire book is possible. This could at least be somewhat helpful to scholars.

Book URL:
http://books.google.com/books?id=5NHZ2LfKj6wC&pg=PP1&dq=psychology+of+the+child&lr=&sig=ACfU3U22gNB2q_4GLh1xJOGALGtNmdFvjQ#PPR5,M1

Date accessed: August 16, 2008

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt **Date of investigation**: 6/28/08

Book title: Second Language Acquisition An Introductory Course (second edition)
Author: Susan M. Gass, Larry Selinker
Year of Publication: 2001
Place of Publication: Lawrence Earlbaum Associates
ISBN number?: 0805835288
Content/Disciplinary area: Applied Linguistics – Second Language Acquisition.

I searched for this book on the GEORGE catalog online, and discovered that there is a more recent, 2008 version. This version is not available for view (not even snippet) on Google Books. When I looked for the book at WorldCat, there were several libraries that own it. There was also a link to an electronic version in NetLibrary! (BOTH 2001 and 2008 versions). It appears I can even chose the language I want to book to appear in. Unfortunately, it gave me the message "This eContent is not included in your library's collection and is unavailable for viewing." (see Image I).

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	Very good condition – Limited Preview from Google Books
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Very. Also available on several OPACs, Georgetown Library. Georgetown's subscription to NetLibrary does not offer access to this book.
MARC record available?	NA
Can the book be searched through an OPAC? If so, which one? **Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)**	No. Other libraries may permit access to this book on their subscription to NetLibrary- Georgetown does not.

SCAN SEARCH

search paradigm settings (example: every 10 double-pages) scanned through entire book

Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
1-16, 21-64, 69-91, 96-111, 116-140, 143-167, 170-191, 194-221, 224-258, 261-309, 312-328, 331-371, 374-397, 400-412, 415-459, 461-462, 464-466, 474-488	these pages are NOT available for view			
	These pages have a message from Google that says: You have either reached a page			

that is unavailable for viewing or reached your viewing limit for this book."
All other pages in this book (that the researcher can view) are in perfect condition – no anomalies.

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	Comment (Is OCR search permitted): Y / N
nativism	8 instances of this term, unfortunately they are all on pages that are not viewable.
Competition Model	successful recognition – many on pages not viewable. When I go to the page (193), the yellow highlighting that Google Books does has not 'reloaded' – it is still highlighted for 'language' (a practice search I had done earlier). The yellow highlighting, which makes the word visible to the researcher, does not remove itself from prior search, and then attach to new word searched. I tried reloading the page – highlighting still did not work.
cues	successful recognition – many on pages not viewable. When I go to the page (193), the yellow highlighting that Google Books does has not 'reloaded' – it is still highlighted for 'language' (a practice search I had done earlier). The yellow highlighting, which makes the word visible to the researcher, does not remove itself from prior search, and then attach to new word searched. I tried reloading the page – highlighting still did not work.
input	same issue as above. The yellow highlighting tool is not working.
interaction	Successful recognition – unfortunately on pages not viewable.
UG	Successful recognition – none of the pages are viewable.
L1; L2	Both successful recognition – none of the pages are viewable.
child acquisition	Successful recognition – none of the pages are viewable.
"Examples in Spanish"	Successful recognition – none of the pages are viewable.
plurals	successful recognition – yellow highlighting worked this time! The word was highlighted in yellow- very easy to find within the text.

IMAGE ASSESSMENT

If the original text includes graphics/illustrations, are they included in the electronic version?	NA – the only graphics in this book are graphs showing data results.
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	
Is the scan on a gray-scale or color scale?	
Quality – does it look like the real thing?	
Illustrations? (would the	

graphics/illustrations be useful to art historians, visual culture researchers?)	
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	
Other notes	

GENERAL OBSERVATIONS

Comment on ease of discovery in integrated interface (OPAC or better)	Very easy
Overall quality of scan of this book	EXCELLENT
Ease of access to full text	No – limited viewing
Ability to copy, quote, cite?	No
Ability to search?	Yes
Ability to download?	No
Ability to print?	No
Scanned book allow for consulting of contents?	No
How successful is OCR recognition?	Excellent – the best I've seen yet
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Definitely
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes.
Issues with this book that would affect usefulness for scholars	None – though yellow highlighting tool could be improved. That would make searching for terms within large bodies of text easier. Highlighter functioning seems to be arbitrary.
Can this book be shared with other scholars/students?	Yes, by emailing URL.
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes

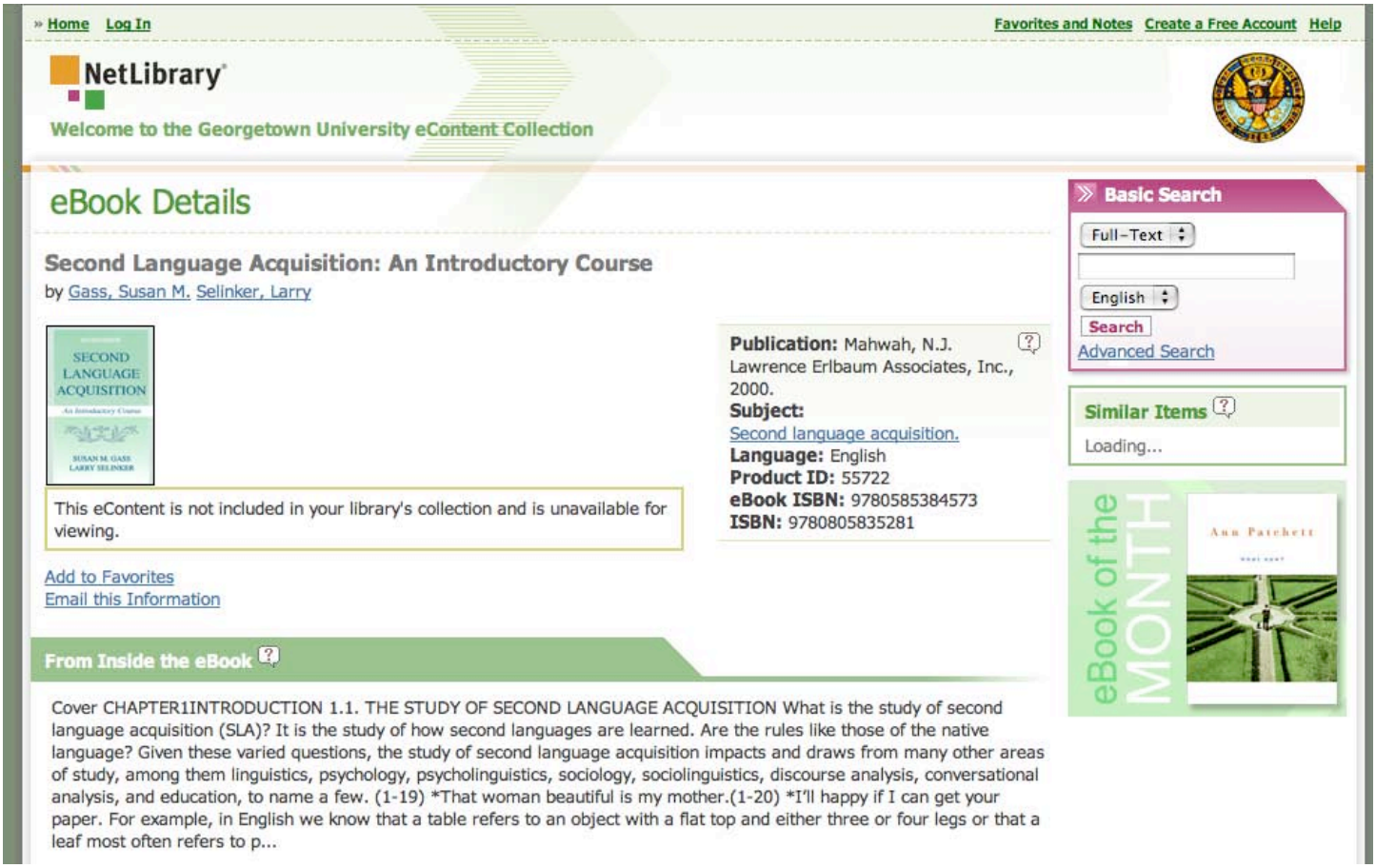
CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	No – although it would be nice if Google Books were to digitize volumes I (1994) and III (2008
Other notes	Fantastic reference book for any applied linguist studying SLA

Book URL: http://books.google.com/books?id=fFw-Ep1wUPgC&pg=PA16&vq=child+acquisition&dq=second+language+acquisition&lr=&source=gbs_search_s&sig=ACfU3U3WOlXbd_01oeOd33t0rYgqDyjNIQ#PPA19,M1

Date accessed: 6/28/2008

Image I



CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt **Date of investigation**: 6/28/2008

Book title: Teaching and Learning Chinese as a Foreign Language A Pedagogical Grammar
Author: Janet Zhiqun Xing
Year of Publication: 2006
Place of Publication: Hong Kong University Press
ISBN number?: 9622097634
Content/Disciplinary area: Language Learning, Chinese as an FL

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	Excellent condition.
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Very easy – no metadata.
MARC record available?	NA
Can the book be searched through an OPAC? If so, which one?	No – it is available in several libraries, including the Library of Congress, and also on WorldCat – but none have digital copies. Google Books is apparently the only one.
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	NA.

SCAN SEARCH

search paradigm settings (example: every 10 double-pages) Scanned through entire book.

Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
1-28 (entire first chapter), 33, 38-57, 60, 65-85, 91-132, 137-164, 169-236, 241-274, 280, 284, 290, 295-318	These pages are not provided for viewing.			
281	This is an appendix – very impressive that Google included it, as it is pedagogic material for potential distribution in class. Almost all in Chinese characters. To follow is common books used to teach Chinese, common reference books,			

common study abroad programs, common websites – all available. The rest of this book (those pages that are available for view) is in fantastic condition – no anomalies.

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	Comment (Is OCR search permitted): Y / N
business Chinese	all instances successful – search results all highlighted in yellow. Fantastic.
tone	all instances successful – search results all highlighted in yellow. Fantastic.
formal	all instances successful – search results all highlighted in yellow. In other books, the highlighter tool was an issue. It is working perfectly in this book.
dialect	all instances successful – search results all highlighted in yellow.
topic [as in topic-comment construction]	all instances successful – search results all highlighted in yellow.
I tried to search for Chinese characters. My mac has tradictional and simplified Chinese character keyboard. In the traditional options, *Hanin*, *Cangjie, Jianyi, Dayi(Pro),* and *Pinyin*. I tried ten keys from each of these styles and not one was recognized. (耳口木廾廾薦燎 are examples). Did same with simple to no avail. I then did some internet searches for Chinese characters that I knew were in this book (the tone system and 'ma' is a classic example – ma, with four different tones, has four different meanings in Chinese. I tried 妈 (mother), 麻 (hemp), 马 (horse), and 骂 (curse).	None of these searches were successful. My immediate concern was that it was a font or style issue. But given that I tried several letters from each keyboard option, and not one worked in this book, I wonder if OCR Chinese character recognition does not work. Here are the copied signs I searched for, followed by an image in the book: 妈 (mother), 麻 (hemp), 马 (horse), 骂 (curse) 麻 30), as illustrated below. This tool helped unravel th Tone 1: high level 55 as in *mā* 妈 "mother" Tone 2: middle rising 35 as in *má* 麻 "hemp" Tone 3: low falling rising 214 as in *mǎ* 马 "horse" Tone 4: high falling 51 as in *mà* 骂 "curse"
声调 (this means tone in Chinese – simplified text)	Not successful. 声调 4.1 Introduction Tone (声调, *shēngdiào*) i
爸 (dad/pa), 爸爸 (daddy/papa)	Not successful. *aoliang* "bea *bàba* 爸爸 I am convinced that my search should be picked up – it is the same form. I do not believe that the OCR recognition can do Chinese characters. I may consult with a Chinese speaker in the future to confirm (see if they can type from keyboard).

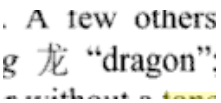
龙 （dragon）	Not successful. (page 89) 
p□ny□n *p□ny□n*	This word (pinyin, for spell sound) is present on page 89, with an i with diacritic on it. I searched for it with the diactric, in regular and in italicized font (in book it is italicized) – search OCR NOT successful.

IMAGE ASSESSMENT

If the original text includes graphics/illustrations, are they included in the electronic version?	NA no images. The front cover could be considered an 'image' – however, none of the text is recognized by the OCR.
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	
Is the scan on a gray-scale or color scale?	
Quality – does it look like the real thing?	
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	
Other notes	

GENERAL OBSERVATIONS

Comment on ease of discovery in integrated interface (OPAC or better)	No OPAC – but is available for partical view in Google Books.
Overall quality of scan of this book	Excellent.

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Ease of access to full text	Not so – only limited pages are available, and text cannot be coppied.
Ability to copy, quote, cite?	No
Ability to search?	Yes – however, Chinese characters are not recognizable, nor are some vowels with certain diactrics (*p□ny□n,* for example).
Ability to download?	No
Ability to print?	No
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	Only English is successful with OCR recognition. Chinese characters are not. This is a significant downfall, given that this book's topic is Chinese grammar and pedagogy.
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Yes.
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Absolutely.
Issues with this book that would affect usefulness for scholars	NO searchability in Chinese text, a significant downfall (note- the Chinese used in this book is simplified).
Can this book be shared with other scholars/students?	Yes, by emailing the link.
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	NA – this is the only volume.
Other notes	For some of the word searches above, I used an online Chinese dictionary, http://www.mandarintools.com/worddict.html. I copied and pasted the Chinese text into the Google Books search tool. I did lots of research on typing Chinese characters, Google options for Chinese characters, downloads, etc. It seems that my mac keyboards are just as sufficient. My only limitation is that I don't know how to write/type Chinese myself.

Book URL:
http://books.google.com/books?id=hTwzUj4ExeIC&pg=PA26&vq=dialect&dq=second+language+acquisition&lr=&source=gbs_search_s&sig=ACfU3U1lgWDRfqX9kE7jC-0E6QWmk_nrqQ#PPA88,M1
Date accessed: 6/28/08

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt **Date of investigation**: Aug. 19, 2008

Book title: The Arabic Language
Author: Kees Versteegh
Year of Publication: 1997
Place of Publication: Columbia University Press
ISBN number?: 023111525
Content/Disciplinary area: Arabic language, historical and literary context, variation and dialect study of Arabic

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	Limited preview book available with Google Books.
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Easily accessible in Google, although limited preview only. Metadata sufficient- book abstract as well as some of table of contents. On Contents page, list of figures and maps are all hyperlinked.
MARC record available?	NA
Can the book be searched through an OPAC? If so, which one?	No
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	Sufficient and accurate. Not all chapters are given in table of contents though.

SCAN SEARCH

search paradigm settings (example: every 10 double-pages)	scanned through entire book

Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
all pages after 25 are not provided for preview.	Also, within the preface pages, a page insert, erroneously labeled 'page 24' is said to not be available for preview. Ordering is incorrect.			
Before page one, an insert, erroneously labeled 'page 28' is said not to be available for preview. Ordering is incorrect.				

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	Comment (Is OCR search permitted): Y / N
Koran / Qur'an	Qur'an successfully found, though in the book is written as Qur'ãn
diglossia	successful detection
al-Andalus	not detected on page 1. Successfully detected on page 255
dialect	successful detection
Aramaic	successful detection
Akkadian	successful detection
Ugarit	successful detection
Ras Samra	successful detection
Rãs Šamra (correct spelling)	not detected – present on page 9
Syriac	successful detection

IMAGE ASSESSMENT

If the original text includes graphics/illustrations, are they included in the electronic version?	There are no images in this the pages available for preview in this text
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	
Is the scan on a gray-scale or color scale?	
Quality – does it look like the real thing?	
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	
Other notes	

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

GENERAL OBSERVATIONS

Comment on ease of discovery in integrated interface (OPAC or better)	Easily accessible in Google Books, not available electronically in an OPAC
Overall quality of scan of this book	Excellent
Ease of access to full text	Poor – limited preview
Ability to copy, quote, cite?	No
Ability to search?	Yes
Ability to download?	No
Ability to print?	No
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	Seemingly good – however one word al-Andalus (above) was not detected successfully. Also, it seems that certain written phonemes are not detected, such as Š, for example.
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Yes, however not all of the pages are provided
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes
Issues with this book that would affect usefulness for scholars	A great source for those wanted to learn/read more about the Arabic language, especially, for example, before taking an Arabic course. Unfortunately the reader does not have access to the entire text.
Can this book be shared with other scholars/students?	Yes, by emailing the URL
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	NA
Other notes	Good, clear map (no color) is provided on page 25. It is of North Arabia before Islam.

Book URL:
http://books.google.com/books?id=2tghviSsrF8C&printsec=frontcover&dq=the+arabic+language&sig=ACfU3U3b6Y3p4sdI7kxZl8_GL9lSjVV4gA#PRA1-PA25,M1

Date accessed: August 19, 2008

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt **Date of investigation**: August 8, 2008

Book title: Tibetan Grammar
Author: Heinrich August Jäschke
Year of Publication: 1888
Place of Publication: Trübner & Company
ISBN number?: na
Content/Disciplinary area: Tibetan language

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	Apparently good condition
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	This book is part of a series of Asiatic and European grammars done by Trübner & Company in the late 1800s; this is the seventh volume. Google Books does not provide any information however on the other volume series. Also, Google's metadata page has some difficulty with certain letters, i.e. "ü", which may be what is making search paradigms for other Trübner grammar series difficult to find. Specifically, the metadata shows this as the publisher: Published by Trub□ner & co., 1888
MARC record available?	NA
Can the book be searched through an OPAC? If so, which one?	No
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	Table of contents is not provided; metadata for this book is limited. However, the table of contents provided in the book has hyperlinks to book sections. This is a great research tool.

SCAN SEARCH

search paradigm settings (example: every 10 double-pages)	searched through entire book

Page(s):	readability/ quality of scan	anomalies?	other?	other?
46, 52, 78 (as examples)		some pages have tips of scanner's fingertips showing in the scan.	This does not prevent readability, but does decrease book's worthiness as potential analog form.	
No other anomalies- pagination in this book is perfect				

WORD SEARCHES DONE WITH THIS BOOK

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	Comment
	(Is OCR search permitted): Y / N
	Note: I very much wanted to attempt to do some searches in the Tibetan language and font, as this would be pertinent to any scholar exploring this language. I tried from several different sources to download the Tibetan font, to include exploring translator and Tibetan language websites. I had no luck! This is a difficult language to be able to work with, from a digital perspective. Frustratingly, my search shall be limited today to English words only.
Lhasa	successful detection, with yellow highlighting
imperative	successful detection. As in previous searches, it takes a while for the highlighter tool to refresh itself and only cover present search paradigm
declension	successful detection
syntax	successful detection
case	successful detection
dialect	successful detection
gender	successful detection
vowels	successful detection
Sanskrit	successful detection (1 instance). However, on page 3, I saw a 'sanscrit' spelling of the word. This was obviously not picked up in the search for 'sanskrit'. The book has both forms. A search for 'sanscrit' yielded several results.

IMAGE ASSESSMENT

If the original text includes graphics/illustrations, are they included in the electronic version?	There are no images in this book.
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	
Is the scan on a gray-scale or color scale?	
Quality – does it look like the real thing?	
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming	

capabilities, filter options)

Other notes

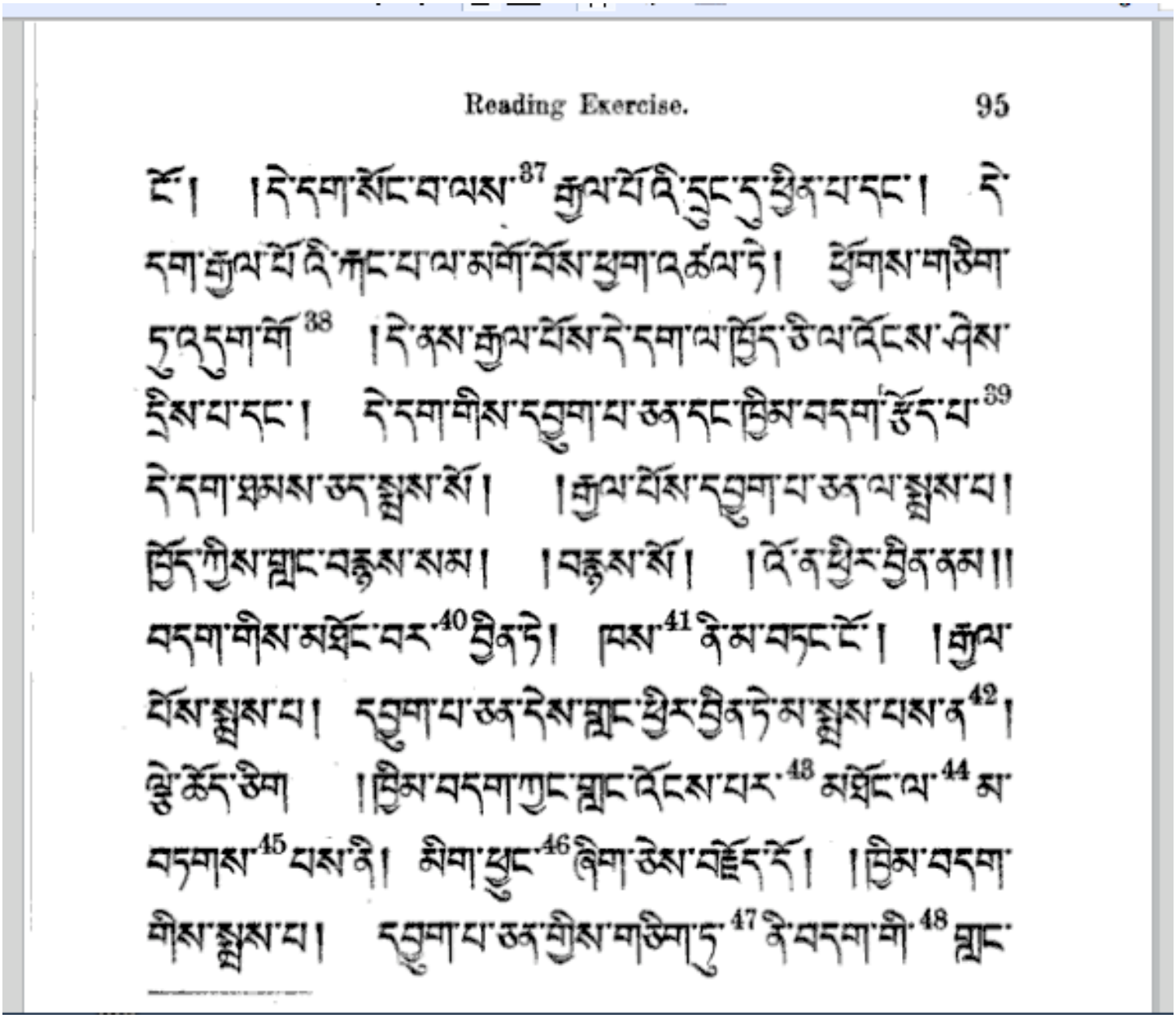
GENERAL OBSERVATIONS

Comment on ease of discovery in integrated interface (OPAC or better)	Easily accessible in Google Books only
Overall quality of scan of this book	excellent
Ease of access to full text	Yes
Ability to copy, quote, cite?	Yes
Ability to search?	Yes, but not for Tibetan letters. This was mostly due to the major difficulty I had in accessing Tibetan fonts online
Ability to download?	Yes
Ability to print?	Yes
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	For English words, 100% accurate
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Yes
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes
Issues with this book that would affect usefulness for scholars	Difficulty in searching for Tibetan words in the digital mode
Can this book be shared with other scholars/students?	Yes, by emailing the URL
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	Metadata is not too detailed; for example, Google does not provide information on the other volumes of this grammar series.
Other notes	

Book URL:
http://books.google.com/books?id=cpIVAAAAYAAJ&printsec=frontcover&dq=tibetan+grammar&lr=&as_brr=1#PPP1,M1

Date accessed: August 8

Figure 1 Example of Tibetan reading exercise provided in this text.



Reading Exercise. 95

ངོ་། །དེ་དག་སོང་བ་ལས་[37] རྒྱལ་པོ་འི་དྲུང་དུ་ཕྱིན་པ་དང་། དེ་
དག་རྒྱལ་པོ་འི་ཀང་པ་ལ་མགོ་བོས་ཕྱག་འཚལ་ཏེ། ཕྱོགས་གཅིག་
ཏུ་འདུག་གོ[38] །དེ་ནས་རྒྱལ་པོས་དེ་དག་ལ་ཁྱོད་ཅི་ལ་འོངས་ཤེས་
དྲིས་པ་དང་། དེ་དག་གིས་དབྱུག་པ་ཅན་དང་ཁྱིམ་བདག་རྩོད་པ་[39]
དེ་དག་ཐམས་ཅད་སྨྲས་སོ། །རྒྱལ་པོས་དབྱུག་པ་ཅན་ལ་སྨྲས་པ།
ཁྱོད་ཀྱིས་གླང་བརྙས་སམ། །བརྙས་སོ། །འོ་ན་ཕྱིར་བྱིན་ནམ།།
བདག་གིས་མཐོང་བར་[40] བྱིན་ཏེ། ཁས་[41] ནི་མ་བཏང་ངོ་། །རྒྱལ་
པོས་སྨྲས་པ། དབྱུག་པ་ཅན་དེས་གླང་ཕྱིར་བྱིན་ཏེ་མ་སྨྲས་པས་ན[42]།
ལྕེ་ཆོད་ཅིག །ཁྱིམ་བདག་ཀྱང་གླང་འོངས་པར་[43] མཐོང་ལ་[44] མ་
བཏགས་[45] པས་ནི། མིག་ཕྱུང་[46] ཞིག་ཅེས་བརྗོད་དོ། །ཁྱིམ་བདག་
གིས་སྨྲས་པ། དབྱུག་པ་ཅན་གྱིས་གཅིག་ཏུ་[47] ནི་བདག་གི་[48] གླང་

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt	**Date of investigation**: August 24

Book title: Un viaje a Bolivia
Author: Alberto Blancas
Year of Publication: 1900
Place of Publication: Santiago de Chile, Imprenta, Litografía y Encuadernación Barcelona
ISBN number?: NA
Content/Disciplinary area: history, Bolivia

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	original provided by Harvard University
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Metadata provided is accurate and sufficient.
MARC record available?	NA
Can the book be searched through an OPAC? If so, which one?	No
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	includes content, popular passages and some book images. Back of the book index is hyperlinked (page 191)

SCAN SEARCH

search paradigm settings (example: every 10 double-pages)	scanned through entire book

Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
137		scanners finger appears in scan. It has been covered by a digitized white mark, but is quite noticeable		
no other major anomalies				

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	Comment
	(Is OCR search permitted): Y / N
Rio Pasaje	successful detection
Jujuy	successful detection
Humahuaca	successful detection
Cochinoca	successful detection
Montellano	successful detection
montaña	successful detection
pagasen	successful detection
deudores	successful detection
LETANIA, LETANÍA	not detected (this word on page 82), is detected with accent even though accent is not present in book
chola	successful detection – also in a picture on page 107

IMAGE ASSESSMENT

If the original text includes graphics/illustrations, are they included in the electronic version?	Yes
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	No, however, the picture on page 33 for example, has a line divide in it where scan quality seems different.
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	I believe OCR, as rest of book is. Unfortunately, searching for images seems impossible – the text for some picture title is not detected for some reason. On page 33 it is not detected, but on 41, yes. Page 50 is also detected successfully.
Is the scan on a gray-scale or color scale?	grey-scale
Quality – does it look like the real thing?	Somewhat
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	Yes
Are the illustrations searchable?	By the picture title, but some (ex. page 33) are not detected.
Ability to print illustrations? (is there a watermark?)	Only by printing out the entire document
Are the images downloadable? (file size, pixel quality)	within the entire document
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	zooming in, searching for images in the text by picture title (for those that work)

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Other notes	

GENERAL OBSERVATIONS	
Comment on ease of discovery in integrated interface (OPAC or better)	Easily accessible in Google Books
Overall quality of scan of this book	good
Ease of access to full text	Yes
Ability to copy, quote, cite?	Yes
Ability to search?	Yes
Ability to download?	yes
Ability to print?	Yes
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	good, except for some picture titles and capitalized words (LETANIA of above, for example)
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Yes
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes
Issues with this book that would affect usefulness for scholars	some OCR recognition issues
Can this book be shared with other scholars/students?	Yes
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	NA
Other notes	(This book was signed by original author on the Title Page – a historic treasure!)

Book URL: http://books.google.com/books?id=H2ICAAAAYAAJ&pg=PA60&dq=arsans&lr=lang_es&as_brr=1#PPT1,M1
Date accessed: August 24

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt **Date of investigation**: July 23, 2008

I am continuing search for early volumes on the Chinese language, particular to pedagogy and Chinese writing system. Herbert Giles wrote several volumes in the late 1800's, almost all of which are *not* available in Google Books (i.e. Synoptical Studies in Chinese Character, Notes on Chinese Composition, Chinese without a Teacher, Chinese Sketches, Chuang Tzu: Mystic, Moralist and Social Reformer, to name a few.) A Dictionary of Colloquial Idioms in the Mandarin Dialect is in fact his only book available for full view in Google Books.

Book title: A Dictionary of Colloquial Idioms in the Mandarin Dialect
Author: Herbert Allen Giles
Year of Publication: 1873
Place of Publication: Shanghai (A. H. De Carvalho)
ISBN number:
Content/Disciplinary area: Chinese, dictionary, colloquialisms

AVAILABILITY SEARCH

Source of scanned book:	Google Books – this book provided by Harvard University
Comments on condition/provenance of book:	This book seems to be rare- no university in the Consortium has it (including Gerogetown). A WorldCat search indicates only 14 universities worldwide own this book.
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Easily found on Google Books- nice that it is available to public in this way. No other editions.
MARC record available?	Na
Can the book be searched through an OPAC? If so, which one?	No
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	No book metadata

SCAN SEARCH

search paradigm settings (example: every 10 double-pages) scanned through entire book. There are some pages before page 1 that are not numbered- these are like a papyrus scroll. Starting at page 1, the researcher can jump from page to page. Google's page 1 corresponds with the book's actual page 1.

Page(s):	readability/ quality of scan	anomalies?	other?	other?
third page		scanners hand, finger gloves and turning of page are all captured in scan! (figure 1)		
3		starting on this page, Google books does not recognize page 3. Page numeration erroneously starts with page 3 when it is really page 5.	Starting on page 7, book's pages and Google's digitization pages correspond again.	
10		Google's pagination		

		recognizes this as page 11		
15		Google's pagination recognizes this as page 16		
35		After page 35, pages 32, 33, 34 and 35 (again) are all repeated, scanned a second time.	Google's pagination just recognizes all of these extra inserts at '35'	
37		After page 37, pages 34, 35, 36, and 37 (again) are all repeated, scanned a second time.	Google's pagination recognizes all of these repeats as '37'	
39		After page 39, pages 36, 37, 38 and 39 (again) are all repeated, scanned a second time.	Google's pagination recognizes these oddly as 40, 43, 44, and 47 respectively.	40 is then erroneously recognized at page 48 by Google's pagination.
41		Google's pagination recognizes as 51. After page 41, pages 38, 39, 40, 41 (again) are all repeated, scanned a second time.	Google's pagination recognizes these repeats as 52, 55, 55 and 41 respectively.	
42		Google's pagination recognizes this page as 42.	43 as 44 and 44 as 44. Starting on page 45, the two pagination orderings correspond once again.	
57		Google's pagination erroneously recognizes this page as 58.		
62	this page is missing			
65		after page 65, pages 63, 64, and 65 (again)	These pages are recognized by Google's pagination as 66, 66 and 67 respectively.	

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	Comment
	(Is OCR search permitted): Y / N OCR recognition in this book is very poor and at random. Most ENGLISH words (let alone Chinese) were not detected at all.
Agree	On page 1, this is one of the dictionary categories. Not detected at all.
Afford	On page 1, this is one of the dictionary categories. Not detected at all.
ko	A spelled out Chinese word, *ko* was detected correctly on pages 37 and 52, but not on 1, 2, among others
Break	On page 4, this is one of the dictionary categories. Not detected at all.
Find, must, do, fell	NONE are detected. Conclude that none of the English words in this book are OCR detectable. Ironically, the Chinese spelled out words are (ko, pu, lo, lai etc.)
shang	Not detected (on page 16)
Keep	Not detected (a dictionary category, page 26)
make, manage	Not detected. Both are dictionary categories, page 32.
pa	Successful detection – a Chinese word spelled out. Ex on page 36.
settle, shake	Not detected. Both are dictionary categories on page 45.
犯人 **(criminal)**	Not detected (on page 63). Conclude that no Chinese letter recognition can be done with OCR recognition in this book.

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

IMAGE ASSESSMENT

If the original text includes graphics/illustrations, are they included in the electronic version?	NO IMAGES IN THIS TEXT
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	
Is the scan on a gray-scale or color scale?	
Quality – does it look like the real thing?	
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	
Other notes	

GENERAL OBSERVATIONS

Comment on ease of discovery in integrated interface (OPAC or better)	Easily accessible in Google Books
Overall quality of scan of this book	Page and text quality (Chinese characters) are all excellent quality
Ease of access to full text	Good
Ability to copy, quote, cite?	Yes
Ability to search?	No – OCR search recognition is EXTREMELY poor in this text. Not only are the Chinese texts not recognizable, most of the *English* words are not detected. Practically no searching can be done with this text.

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

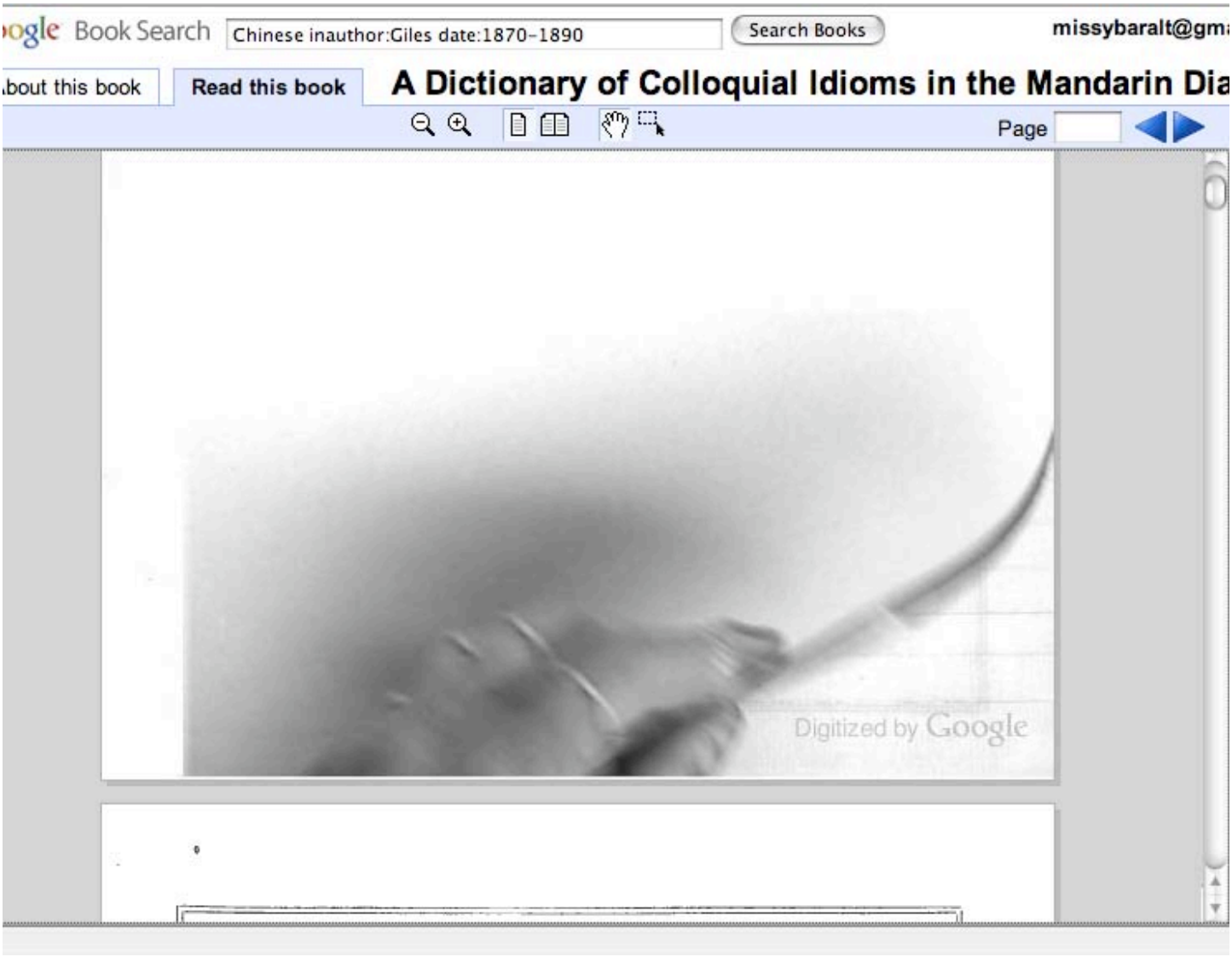
Ability to download?	Yes
Ability to print?	Yes, if book is downloaded
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	Very poor- unsuccessful
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Aside from the third page, which shows scanner's entire hand (Figure 1), this book would qualify as an archive, once pagination problems are fixed.
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Absolutely.
Issues with this book that would affect usefulness for scholars	No searchability in this book. Pagination is completely off – this would cause a problem in being able to search manually in this book, as several pages are repeated and listed erroneously as other page numbers that they're not.
Can this book be shared with other scholars/students?	Yes.
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes.
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	
Other notes	The main problem with this text, besides not being searchable, is the massive pagination errors. A total of 20 pages are scanned a second- some, even a third time – and repeated in this book. This sets off the pagination terribly.

Book URL: http://books.google.com/books?id=kq2uPwfisSMC&pg=RA2-PA70&dq=Chinese+inauthor:Giles+date:1870-1890&lr=&as_brr=0#PRA2-PT1,M1

Date accessed: July 23, 2008

Figure 1

The third page of this scanned book:



CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt **Date of investigation**: July 31, 2008

Book title: The Principles of Teaching Based on Psychology
Author: Edward L. Thorndike
Year of Publication: 1916
Place of Publication: A. G. Seiler, New York
ISBN number?: na
Content/Disciplinary area: pedagogy, psychology

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	Seemingly good condition – this book was written for Teachers College of Columbia University. Book belongs to Harvard University.
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Very easy access within Google Books
MARC record available?	NA
Can the book be searched through an OPAC? If so, which one?	No
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	very nice hyperlinked chapter sections in the Table of Contents. Upon clicking on hyperlink, directs reader to page.

SCAN SEARCH

search paradigm settings (example: every 10 double-pages) scrolled through entire book

Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
71, 72, 171, 217	these pages have fantastic figures/ graphs. Apparently hand-drawn with typed numbers.			
82	hand written algebraic equations on this page- great quality.	None of the letters/numbers on this page are detected by OCR		
285-293	great hyperlinked page numbers (see Figure 1)			

no anomalies in this book – pagination

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

OCR: YES

Word	Comment
attention	successful recognition
individual differences	successful recognition
memory	successful recognition
deductive reasoning	successful recognition
stimuli	successful recognition
motor skill	successful recognition
psychology	successful recognition
theory	successful recognition
immigrant / immigrants	successful recognition
imitation	successful recognition
capacity	successful recognition
behavior	successful recognition
task	successful recognition

The OCR capability in this book is perfect. No problems, except for the hand-written notes in this book.

IMAGE ASSESSMENT

If the original text includes graphics/illustrations, are they included in the electronic version?	No images in this book
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	
Is the scan on a gray-scale or color scale?	
Quality – does it look like the real thing?	
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	
Other notes	

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

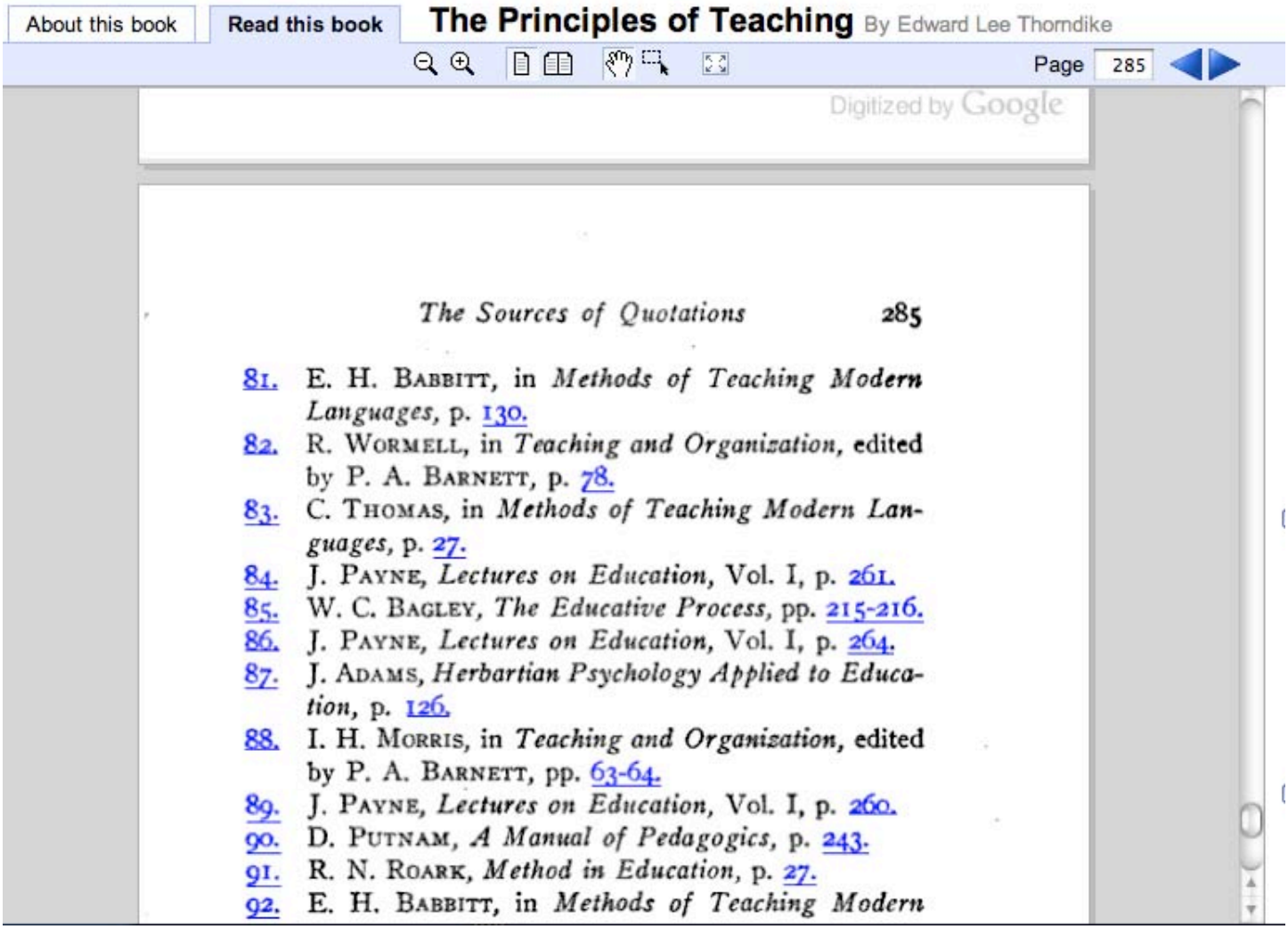
GENERAL OBSERVATIONS

Comment on ease of discovery in integrated interface (OPAC or better)	Easily accessible in Google Books
Overall quality of scan of this book	Excellent
Ease of access to full text	Excellent
Ability to copy, quote, cite?	Yes
Ability to search?	Yes
Ability to download?	Yes
Ability to print?	Yes
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	100% for typed text – perfect
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Absolutely
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes
Issues with this book that would affect usefulness for scholars	None – this is a nice reference to have available online to scholars, and it mentions several constructs pertinent to the field of Applied Linguistics – and how they were described/considered 100 years ago. Very interesting.
Can this book be shared with other scholars/students?	Yes – by emailing the URL
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	NA
Other notes	This book provides suggested activities for teachers and also suggestions for further reading. Page 68 provides one of the first descriptions I have ever seen of 'individual differences' – an important construct today in the field of Applied Linguistics and SLA. Page 106 discusses the construct of attention.

Book URL: http://books.google.com/books?id=pyQBAAAAYAAJ&printsec=frontcover&dq=subject:%22teaching%22+date:1800-1920&lr=&as_brr=1

Date accessed: July 31

Figure 1 example of metadata in this book (hyperlinked page numbers)



CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt **Date of investigation**: August 4, 2008

Book title: Prisciani Grammatici de Laude Immperatoris Anastasii, et de Ponderibus et Mesuris Carmina
Author: Priscian
Year of Publication: this edition was published 1828
Place of Publication: P. J. Schalbacher et socios
ISBN number?:
Content/Disciplinary area: measurement

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	provided by Harvard University- apparently good quality
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Surprisingly very little is provided by Google on this text's metadata. For example, does not state place of publication – only publishers
MARC record available?	NA
Can the book be searched through an OPAC? If so, which one?	No
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	Unfortunately, metadata unavailable

SCAN SEARCH

search paradigm settings (example: every 10 double-pages) searched through entire book

Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
xii	This page has some symbols (perhaps hand written?) that are not detected (example, r with tilde over it)			
xiv		Scanner allowed for pages to move during scan; they appear on this page		
13		page corner was flipped over (dogtagged) during scan, thus covering some of the text		
110			very tip of scanner's finger appears – a minor anomaly	This also happens on pages 143 and 144, however these are end of book.

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	Comment (Is OCR search permitted): Y / N
Fol.	successful detection
mihi	successful detection
□, □, □	No detection of Greek characters (see for example pages 24, 32, 43)
Omnes	successful detection
familiae	successful detection
confligere	successful detection
manichaea, MANICHAEA	not successful – this word written vertically in a graph on page 77. OCR does not detect it or others.
Ponderibus	Not detected! Surprising, as this word is detected perfectly on other pages (not detected on page 82).
obulus	successful detection
Hispalensem	successful detection

IMAGE ASSESSMENT

If the original text includes graphics/illustrations, are they included in the electronic version?	No images in this book
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	
Is the scan on a gray-scale or color scale?	
Quality – does it look like the real thing?	
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	
Other notes	

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

GENERAL OBSERVATIONS

Comment on ease of discovery in integrated interface (OPAC or better)	Very easily accessible in Google Books
Overall quality of scan of this book	overall excellent
Ease of access to full text	ideal – this text can be downloaded too
Ability to copy, quote, cite?	Yes
Ability to search?	Yes – though not any of the Greek characters, and much of the book is written in Greek. This is an issue.
Ability to download?	Yes
Ability to print?	Yes
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	Almost 100% with roman characters, Latin; NO recognition of any Greek characters
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Yes, though I'm picky and would like for any scanners' fingers to be taken out of scan first
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes
Issues with this book that would affect usefulness for scholars	NO searching and OCR detection of Greek!
Can this book be shared with other scholars/students?	Yes, by emailing the URL
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	Yes
Other notes	

Book URL: http://books.google.com/books?id=bC2Z2Usa4bEC&printsec=frontcover&dq=inauthor:Priscian&lr=&as_brr=0

Date accessed: August 4, 2008

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt **Date of investigation**: 6/26/2008

Book title: Sedulii Paschale cum Aelii cum Antonii Nebrissensis interpretatione nuper
Author: Celio Sedulio, Sancho de Nebrija
Year of Publication: 1553
Place of Publication: APVD INCLYTAM GRANATAM
ISBN number?:
Content/Disciplinary area: Latin

AVAILABILITY SEARCH

Source of scanned book:	Google Books – full view!
Comments on condition/provenance of book:	This is a scan of the original from Complutense University of Madrid. There are hand-written notes on the first page of this book (someone had marked in the original). Also, there is an image drawing of what appears to be a hand with a finger pointing (see Image 1). This image is on the f irst three pages of the book. I'm not sure if it was also drawn by hand by some past patron… This image is on the next two pages.
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Good.
MARC record available?	na
Can the book be searched through an OPAC? If so, which one?	No.
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	None.

SCAN SEARCH

search paradigm settings (example: every 10 double-pages)	Searched through entire book. One significant issue with this Google book is the page system. After page i, there is no pagination. This prevents me from jumping to page to page, and also from knowing what page I'm on. I will have to come up with my own pagination system - after page i, I will call the following page 1, then 2, then 3, etc.

Page(s):	readability/ quality of scan	anomalies?	other?	other?
13, top line says LIBER	readability is poor in upper right corner/ scan cuts wording off.			
53, top line says LIBER	scan cuts off words on right side			
77, bottom corner says ERGO VIBI	scan is very light, text slight blurred in bottom right corner			
entire rest of book	quality quite good.	This book scan has no significant anomalies. Lightness/blurriness are the same that would be associated with any original text from the		

CLIR Mass Digitization Evaluation Project
Book Search Form

1500s

Word Searches Done with This Book

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	Comment
	(Is OCR search permitted): Y / N **note**: because there is no pagination with this book (a Google Books error), successful searches can send the researcher to the word (and will jump to that page), but the researcher has no idea what the page number is on which it appears.
monumenta	Some successful, some not. I unfortunately cannot list page numbers here.
vilius	Not successful.
poetæ	Not successful.
Athenis	Not successful.
mentes	Gave me one search result, but it could not take me to the word. Word appears in search results, but cannot be shown (or highlighted in yellow). Not successful.
tricabilis	Not successful.
ſemper	Successful, but only with a few (I saw several instances of this word that it did not read). It gave me one result hit, as "/emper." When I went to highlighted word, only 'emper' (ſemper) was highlighted. Difficulty in reading this symbol.
literis	Not successful.
epigramma	Not successful.
cardine	Did not give me the one I found. One successful result, but as with **mentes,** could not take me to the word. Not successful.

Searchability in this book is very poor. This is probably due to the fact that it is an extremely old text and characters are hard to read.

Image Assessment

If the original text includes graphics/illustrations, are they included in the electronic version?	There is one image in this book, on page 103. (see Image II)
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	None
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	OCR – recognition of text items in image unsuccessful.
Is the scan on a gray-scale or color scale?	not sure
Quality – does it look like the real thing?	yes
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	NA
Are the illustrations searchable?	No
Ability to print illustrations? (is there a watermark?)	Yes, printing (and downloading) of entire text available. No watermark.
Are the images downloadable? (file	within the entire document.

size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	Yes- I can zoom. But this would be no different than getting a magnifying class and using it with the original.
Other notes	

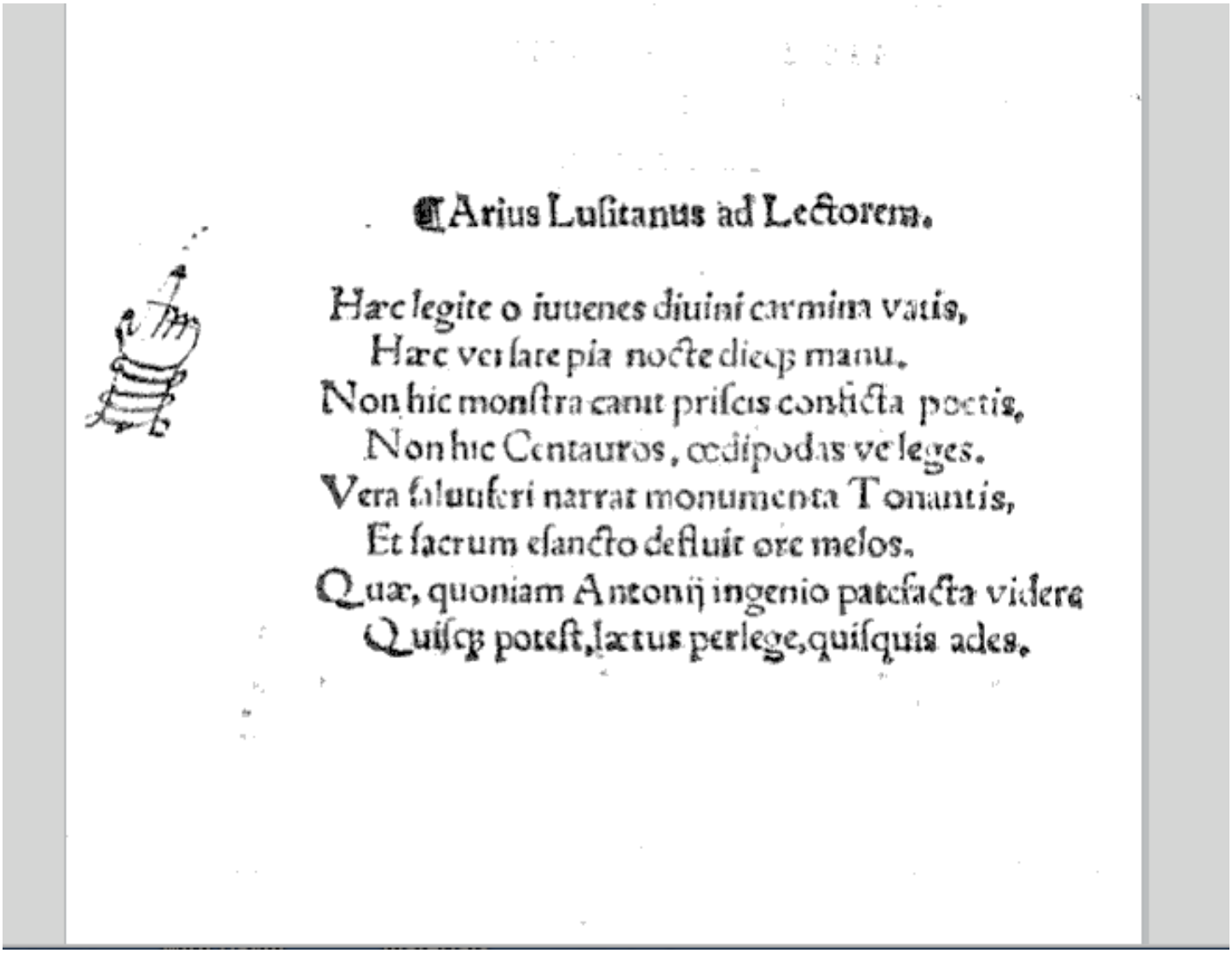
GENERAL OBSERVATIONS

Comment on ease of discovery in integrated interface (OPAC or better)	Not available in OPAC – very easy to access on Google Books.
Overall quality of scan of this book	Very good, for a book from the year 1553.
Ease of access to full text	Good.
Ability to copy, quote, cite?	No.
Ability to search?	I would say no – search ability is extremely poor in this book.
Ability to download?	Yes!
Ability to print?	Yes, once you download book.
Scanned book allow for consulting of contents?	Barely.
How successful is OCR recognition?	Very poor- not successful.
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Absolutely. However, I'm not sure about the three drawings on the first three pages? This may have been done by a library patron.
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes, definitely.
Issues with this book that would affect usefulness for scholars	Pagination not available—I'm not sure why Google didn't do that with this book. Word searches almost impossible- it'd be just like using an analog form.
Can this book be shared with other scholars/students?	Yes.
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes.
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	NA.
Other notes	

Book URL:
http://books.google.com/books?id=C-o8VULcBnYC&printsec=frontcover&dq=inauthor:%22Antonio+de+Nebrija%22&lr=#PPR1,M1

Date accessed: 6/26/08

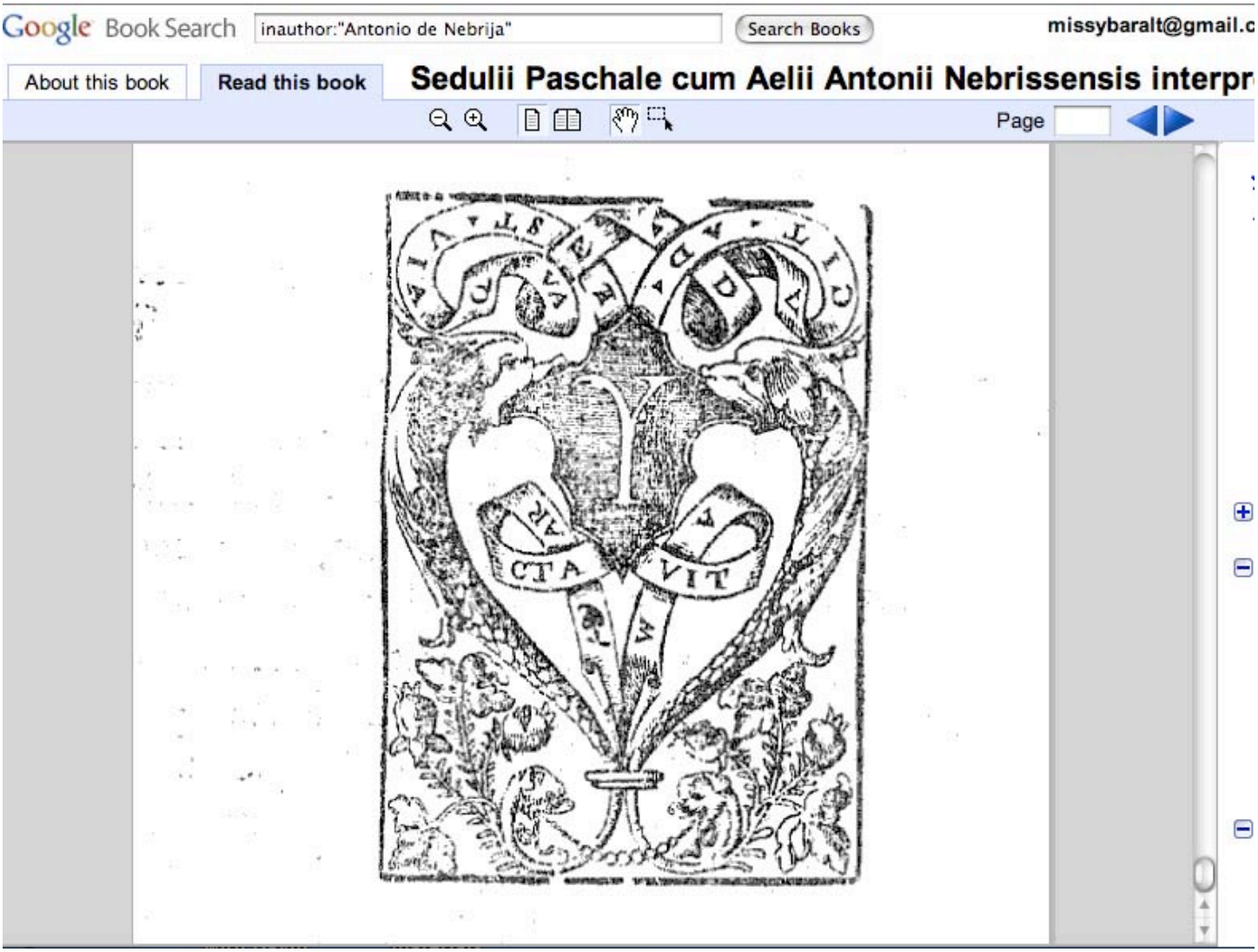
Image I – drawing by patron?



CArius Luſitanus ad Lectorem.

Hæc legite o iuuenes diuini carmina vatis,
Hæc verſate pia nocte dieq; manu.
Non hic monſtra canit priſcis conficta poetis,
Non hic Centauros, œdipodas ve leges.
Vera ſaluiferi narrat monumenta Tonantis,
Et ſacrum ſancto defluit ore melos.
Quæ, quoniam Antonij ingenio patefacta videre
Quiſq; poteſt, lætus perlege, quiſquis ades.

Image II – the only image in this book



CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt **Date of investigation**: July 24, 2008

Book title: Sociocultural Theory and Second Language Learning
Author: James Lantolf (Ed.)
Year of Publication: 2000
Place of Publication: Oxford, Oxford University Press
ISBN number?: 0194421600
Content/Disciplinary area: linguistics, SLA, Sociocultural Theory, Vygotsky

AVAILABILITY SEARCH

Source of scanned book:	Google Books
Comments on condition/provenance of book:	Limited preview on Google books, seemingly excellent condition
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Very easy to find in Google books, although it does not allow full view.
MARC record available?	NA
Can the book be searched through an OPAC? If so, which one?	No – surprisingly, Georgetown or any of the Consortium universities do not own this book.
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	NA

SCAN SEARCH

search paradigm settings (example: every 10 double-pages) scrolled through entire book

Page(s):	**readability/ quality of scan**	**anomalies?**	**other?**	**other?**
5-6, 12-13, 19-20, 26-27, 33-34		These pages are left out		
all pages after 39 are excluded!	Otherwise, entire content / scan quality of this book is excellent.			

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	**Comment** (Is OCR search permitted): Y / N

ZPD (Zone of Proximal Development)	successful detection. Yielded several results, highlighted them all in yellow.
imitation	successful detection, yielded several results, however highlighting is still on past search.
inner speech	successful detection, yielded several results, however highlighting is still on past search.
Vygotsky	successful detection, yielded several results, however highlighting is still on past search. Can't seem to re-do highlighter tool, Vygotsky is not highlighted.
scaffolding	successful detection. Yielded several results, highlighted them all in yellow.
EFL	successful detection, yielded several results, however highlighting is still on past search.
language teaching	successful detection. Yielded several results, highlighted them all in yellow. However, sometimes only gave me 'language' and only highlighted that word. Did not do search with two words together.
"language teaching"	same as above- made no difference.
externalization	successful detection. Yielded several results, highlighted them all in yellow.
mediation	successful detection. Yielded several results, highlighted them all in yellow.

IMAGE ASSESSMENT

If the original text includes graphics/illustrations, are they included in the electronic version?	No images in this book (or at least in the pages available for preview)
If there are no graphics/illustrations are there other visual elements missing? (wormholes, ink smears, page aging/mold, paper quality)	
Is the image scan an OCR or a PDF? If OCR, how successful is the recognition?	
Is the scan on a gray-scale or color scale?	
Quality – does it look like the real thing?	
Illustrations? (would the graphics/illustrations be useful to art historians, visual culture researchers?)	
Are the illustrations searchable?	
Ability to print illustrations? (is there a watermark?)	
Are the images downloadable? (file size, pixel quality)	
Does the electronic version allow for any advantages that the print version would not allow? (zooming capabilities, filter options)	
Other notes	

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

GENERAL OBSERVATIONS

Comment on ease of discovery in integrated interface (OPAC or better)	Accessible in Google Books only, though for only very limited preview.
Overall quality of scan of this book	Excellent – perfect.
Ease of access to full text	No
Ability to copy, quote, cite?	No
Ability to search?	Yes- search and OCR recognition is fantastic
Ability to download?	No
Ability to print?	No
Scanned book allow for consulting of contents?	Yes
How successful is OCR recognition?	Perfect- 100% accurate, with results highlighted in yellow. It sometimes takes a while though for highlighting tool to renew itself to the new searched word.
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Yes
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes
Issues with this book that would affect usefulness for scholars	None – though this is a limited preview book.
Can this book be shared with other scholars/students?	Yes, by emailing the URL
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	Yes
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	NA
Other notes	Fantastic quality and searchability in this book.

Book URL:
http://books.google.com/books?id=imwsewtZKSMC&printsec=frontcover&dq=lantolf&sig=ACfU3U1D81WQmFioNrUAqrOYWtPRbpLnvQ
Date accessed: July 24, 2008

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Researcher name: Melissa Baralt **Date of investigation**: 6/15/08

Book title: Logical Structure of Linguistic Theory
Author: Noam Chomsky
Year of Publication: 1955
Place of Publication: MIT
ISBN number?: none, does have Order Number: 91920
Content/Disciplinary area: Linguistics

AVAILABILITY SEARCH

Source of scanned book:	MIT. This is a complete digitized copy of the thesis draft, including Chomsky's handwritten annotations, provided generously to the public by MIT. Website is: http://alpha-leonis.lids.mit.edu/chomsky/. Original format is microfiche. According to the above site, this document is available in limited form from ProQuest.
Comments on condition/provenance of book:	very poor. First few pages have a strip running down right side that white out many of the letters.
Ease of finding book, appropriateness of available metadata in the project supplying book (i.e. multivolume works, books with different editions/dates/titles)	Found it through internet search. Metadata NA.
MARC record available?	No
Can the book be searched through an OPAC? If so, which one?	ProQuest
Comments on accuracy/ usefulness of book metadata (i.e. table of contents? guide to book?)	NA

SCAN SEARCH

search paradigm settings (example: every 10 double-pages) every 10 pages

Page(s):	readability/ quality of scan	anomalies?	other?	other?
4	poor	blurriness, spotting	some words are illegible	
14	ok	some words have blurred ink	neat to see Chomsky's personal notes on this page	
24	good			
34	good			
44	good		personal notes	
54	ok	lightness		
64	ok	lightness		
74	good	black splotches at bottom of page	personal notes	
84	ok	black splotches at bottom of page, darkness on right side	lots of personal notes on this page	
94	good	black splotches	personal notes	
104	good		personal notes	

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

114	good	black splotches	hand written syntactic derivations (type writer impossible back then?)
124	good	black splotches	lots of personal notes, cross out of text
134	good	black splotches	personal notes
144	ok		personal notes
154	good	black splotches	personal notes

WORD SEARCHES DONE WITH THIS BOOK

(at least 10, include footnotes, endnotes, comment on accuracy of finding hits on pages I've already examined above)

Word	Comment
	(Is OCR search permitted): Y / N No OCR searching permitted in this PDF document

GENERAL OBSERVATIONS

Comment on ease of discovery in integrated interface (OPAC or better)	Limited access in ProQuest. This document made available by MIT website.
Overall quality of scan of this book	Ok
Ease of access to full text	Excellent
Ability to copy, quote, cite?	No
Ability to search?	No
Ability to download?	Yes
Ability to print?	Yes
Scanned book allow for consulting of contents?	No
How successful is OCR recognition?	No OCR recognition
Worthy as archive? (i.e., would this copy be acceptable as an archival copy of the book if physical copies were lost or destroyed?	Yes, though some pages have very poor scan quality. (This is basically a scanned copy of analog version)
Reproduction – is quality of scan good enough that this book could be reproduced into analog form?	Yes – the personal notes by Chomsky make this a timeless edition
Issues with this book that would affect usefulness for scholars	No searches possible

CLIR MASS DIGITIZATION EVALUATION PROJECT
BOOK SEARCH FORM

Can this book be shared with other scholars/students?	Yes
"web 2.0 functionality" – can digitized text be reprocessed in order to support other kinds of research/ use?	No – only PDF pages
Predictability and reliability of collection – any items missing (i.e. volumes in a set, standard words, etc.)	None
Other notes	Nice to have this document made available to public. However it does not permit for any searching.