

# Transcribing Scientific Textbooks for Secondary School Blind Students the Italian Experience

Cristian Bernareggi  
Università degli Studi di Milano  
Milan, Italy  
cristian.bernareggi@unimi.it

Giuseppe Nicotra  
ArchiBraille Association  
Verona, Italy  
gn@lambdaproject.org

## Abstract

Making available scientific textbooks to blind students on time for their learning tasks at school (i.e. lessons, homework, exams, etc.) has been a challenge for a long time. In recent years, computer applications to assist blind students in writing, reading and processing mathematical expressions with a Braille display and a speech synthesizer have been developed and, at present, they are used by many blind students (e.g. the LAMBDA system, ChattyInfty, etc.). Furthermore, the development of InftyReader, the most advanced OCR program in the world for mathematical formulae, has made possible a new workflow for the production of scientific textbooks for blind students: the formulae are not transcribed character by character, but they are automatically recognized by InftyReader and corrected by the transcriber. This workflow may highly speed up the transcription process. In this paper we present the Italian experience in the production of scientific textbooks for blind students in primary and secondary school by using InftyReader to recognize mathematical formulae, Infty2Braille to import the resulting document into the LAMBDA system, and the LAMBDA system to enable blind students to read the textbook.

## 1 Introduction

Enabling blind students to study mathematics from the primary school to university is a well-known challenge. On the one hand, blind students need tools to read, write and process mathematical expressions and diagrammatic representations widely used in scientific communication. On the other hand, blind students need study material (i.e. textbooks, exams, notes, etc.) in accessible formats (e.g. in Braille on paper, as tactile representations, as audiobooks, etc.) on time for their learning activities at school.

The former issue has been tackled for years and nowadays there are many assistive tools which can be used by blind students to read, write and process mathematical expressions. Actually, in recent years, the advantages brought by the use of computer with Braille display and speech synthesizer in many literary subjects, have been appreciated also in scientific subjects. In particular, in Italy, the LAMBDA system [1] has been getting widespread among blind students in primary and secondary school.

The LAMBDA system is a computer application which enables the student to read, write and process scientific documents containing mathematical formulae. This program makes available a wide variety of features to explore and edit mathematical formulae (e.g. hierarchical exploration of nested expressions, select, cut and paste of mathematical elements, editing of tables, etc.) as well as to improve communication between sighted assistants and blind students (e.g., sequential and bidimensional representation of formulae).

The latter issue, concerning the efficient and effective production of accessible scientific reading material for blind students, has been addressed in many ways, partly based on transcription tools for Braille (e.g. Duxbury programs), partly recording audiobooks. Nonetheless, at present, in Italy, there are still very few scientific textbooks available for blind students.

Some reasons account for this lack of scientific documentation:

- the transcription workflow from the text on paper into Braille or into an audiobook requires high expertise. Because of the complexity of mathematical notation, the transcriber or the reader must know both Braille mathematical notation and mainstream mathematical notation in depth. In many situations, also knowledge of the subject is crucial to succeed in producing study material really usable by blind students. Therefore, transcription centres run into the difficulty to find expert transcribers or readers and to keep prices low enough for the end user. Furthermore, students must ask the transcription centres for a book to be transcribed, many months in advance;
- in the process of transcription of scientific textbooks, OCR applications are not used as it happens with literary material. Therefore, the number of hours needed to process a textbook is considerably higher than with literary content. The transcription time for one page of a scientific textbook depends on the structure of the page and on the number of formulae. Based on the workflow used in many transcription centres in Italy, one page of a secondary school mathematics textbook can be properly transcribed in one hour and half on average. This affects both the price and the production time.

Because of these reasons, in Italy, scientific textbooks for secondary schools are very seldom transcribed into Braille or recorded as audiobooks. Support teachers, volunteers and often members of the family of blind students try to compensate for the lack of scientific learning material. Despite the great efforts, blind students do not have scientific textbooks, notes and exercises on time for their learning tasks. Consequently, blind students in secondary schools experience many disadvantages due to the lack of learning material. Many students are advised by teachers themselves to follow special programs for mathematics and not to undertake scientific studies after secondary school.

Based on these reasons, in Italy, a project has been undertaken for 2008 in order to go towards an efficient and effective transcription service of scientific material for students in secondary school. It was divided into two projects: the Infty2Braille project [2], supported by the Italian ministry of education and the Toccare i Numeri (i.e., Tactile Numbers) project. The Infty2Braille project, which ended in 2009, developed the Infty2Braille application, which automatically converts a PDF document containing mathematical formulae into a document usable in the LAMBDA application. This module is based on the InftyReader OCR [3] which recognizes mathematical expressions and converts them into MathML. The Toccare i Numeri project aimed at producing accessible scientific textbooks, usable in the LAMBDA system, for primary and secondary school students in four Italian regions. To this purpose, a novel transcription service of scientific material was set up. This service is based on the use of InftyReader as OCR program, on Infty2Braille and on ChattyInfty for intermediate editing. In the project were involved provincial administrations which are in charge of providing schools with textbooks for blind students and the Italian Union of the Blind (Ancona) to facilitate the communication with blind students and schools. The service was set up and maintained by the association ArchiBraille (Verona).

## 2 Transcribing scientific educational material in Italy

In Italy, blind students in primary and secondary school are provided with textbooks according to the following stages. The provincial administration is in charge of making available textbooks to blind students in primary and secondary school. Therefore, first of all, the family of the blind student sends the provincial administration a copy of the textbooks to be used at school the upcoming year. This list of books is usually made available by the schools in March. Based on instructions provided by school teachers, not the whole textbooks are required to be transcribed into an accessible format, but only the parts to be actually used. Both transcription into Braille on paper or in an accessible electronic format can be required to be delivered.

Subsequently, the provincial administration asks a specialized transcription centre for a quotation. Once the provincial administration has accepted the quotation, the transcription work can start (approximately in June). It is worth noting that in Italy, over the 90% of the requests of transcription of scientific textbooks is not accepted by provincial administrations. This is due to two main factors: the quotation for the transcription into Braille of a scientific textbook is often too high to be accepted (over 3000 EUR for a 250 pages textbook on average); the estimated transcription time is not acceptable for the student needs (e.g. 8-12 months). Because of these problems, blind students in primary and secondary schools usually adopt alternative strategies to read scientific textbooks. The most common solution is to ask a sighted assistant to read them only the parts strictly needed, after the lesson. This solution is very time consuming for the student who has to transcribe him/herself the notes, it is error-prone because the sighted assistant is not always expert in mathematics and it leads to not reusable learning material.

Another alternative solution consists in studying on old textbooks already available in Braille. This solution may lead the student to attend a program different from his/her peers.

In the end, one more solution consists in requiring audiobooks instead of electronic or Braille textbooks. Unfortunately, even for audiobooks transcription time and prices are high, and, especially for mathematics, students often run into difficulty while listening to formulae.

### 3 The Transcription Service for Scientific Textbooks

The transcription service was set up taking into account the need for quick transcription of scientific textbooks. To this purpose, all the stages of the transcription process were optimized.

First of all, thanks to communication actions towards the schools and provincial administrations, the list of scientific textbooks is received in March and for each textbook the teacher specifies which parts are to be used and in which months of the school year. This makes it possible an optimal scheduling of the work so that all students can receive their learning material in due time, not all at the beginning of the school year.

Subsequently, the textbook in PDF format is required to the publisher to be processed. This is possible thanks to agreements with publishers which agree to deliver PDF files to transcription centres for the blind. Some publishers (e.g. Zanichelli which is one of the main publishers of textbooks for primary and secondary schools) are able to deliver the PDF file in a very short time (i.e., some minutes).

The subsequent step concerns the actual transcription process. The PDF file is processed through InftyReader which extracts text from the PDF file and recognizes formulae. Subsequently, the document is analysed by the transcriber and many errors are corrected by using ChattyInfty. Once all the corrections are made, the document is imported in the LAMBDA system through the Infty2Braille module. Once in the LAMBDA system, the transcriber checks all the mathematical formulae for errors. In the LAMBDA system, formulae are represented in a sequential notation which is readable by a sighted user. This feature is of help also for the transcriber who can read the formulae and correct mistakes even if she/he does not know the Braille rules for mathematics. Furthermore, the transcriber writes alternative text descriptions for figures or provides references to tactile representations when available.

### 4 Results

This transcription service started in March 2011. Up to October 2011, 4500 pages of primary and secondary school textbooks of mathematics have been transcribed. The average time for transcribing a page is approximately 20 minutes on average. This result is encouraging if compared to the average time for traditional transcription.

In October 2011 the students and teachers who received the transcribed material were interviewed in order to improve the service. They remarked the following points:

- students mostly use exercises with solutions and homework transcribed in LAMBDA. They find very helping the use of LAMBDA system for solving the exercises, because they can select, copy, cut and paste expressions or parts of them;
- the instructions given by teachers concerning the scheduling of the topics presented in the program are precisely followed and no rescheduling of transcription has been needed up to now;
- two users needed a different layout for the text in the LAMBDA editor in order to be facilitated in reading with their Braille display. This could be achieved straightforwardly;
- teachers appreciate the use of the LAMBDA system especially for exercises in the textbooks because they can add comments and notes useful for the student, where needed in the document.

## References

- [1] Giuseppe Nicotra, Giovanni Bertoni, Enrico Bortolazzi, and Luciana Formenti. Innovative methods for accessing mathematics by visually impaired users. In *Proceedings of the 12th international conference on Computers helping people with special needs*, ICCHP'10, pages 423–430, Berlin, Heidelberg, 2010. Springer-Verlag.
- [2] Infty2braille project. [http://www.copernico.bo.it/index.php?option=com\\_content&view=article&id=277&Itemid=294](http://www.copernico.bo.it/index.php?option=com_content&view=article&id=277&Itemid=294).
- [3] Infty project. <http://www.inftyproject.org>.