

# Comparing Text Editions with the Aid of the Computer<sup>1</sup>

Ben J. P. Salemans

UCI, University of Nijmegen, Geert Grooteplein 41, NL-6525 GA Nijmegen, The Netherlands

e-mail: B.Salemans@uci.kun.nl or B.Salemans@let.kun.nl

**Key words:** automation, bibliography, book history, edition, fingerprint, identification, scanning, signature, STCN

## Abstract

The aim of this paper is to reveal the limitations of relying on “fingerprint formulas” or “signature positions” to identify printed editions, in the manner of the STCN project. The paper proposes that scanned images of texts be used to identify editions.

## Introduction

Many bibliographers and book historians try to solve the question of how differences between two or more printed editions of the same work can be recognized and described with precision. When two editions mention exactly the same date of printing on the title pages or in the colophons, this does not necessarily imply that they have been printed with the same setting of printing types. Logically, small differences in the typography of an edition can offer us information on the date of printing and the relationship between different editions. “Fingerprint formulas” (synonyms: “fingerprints,” “signature positions”) of imprinted text editions are used to discern such differences.

In this paper the use and development of fingerprints will be explained and criticized. Because fingerprint formulas can be an inaccurate method of identifying an edition, there is a need for additional or different information. This paper proposes that scanned images of texts be used to identify editions. These scans can be manipulated and investigated with

the aid of the computer in several ways. Although this paper will concentrate on fingerprints, it will not be difficult to envision other applications. If scans can be used by book researchers to compare fragments of editions, so, for instance, art historians can use them to compare Rembrandt etchings with forgeries. Furthermore, researchers dealing with small, ancient text fragments found in old bindings can, once these fragments have been scanned, cut, paste and manipulate them with ease on the computer screen.

## Description of Fingerprint Formulas

The “fingerprint formula” or “signature position” to identify editions is a well-known procedure, recommended in the reference works of Foxon (1970), Gaskell (1979, p. 333) and Laufer (1972, p. 137). In the Netherlands the fingerprint formula is used in the prestigious Short-Title Catalogue Netherlands (STCN) project, which aims to produce a retrospective national bibliography for the period 1541–1800. In Vriesema (1986a,b) and Handleiding STCN (1988, § 83–87, pp. 65–68) there are descriptions of how a fingerprint formula has to be developed.

A fingerprint formula consists of two basic parts. The first part mentions the year in which the edition appeared, plus the book size; “04” stands for quarto size, “08” for octavo size, etc. Therefore, “156008” as the first part of a fingerprint formula represents an

---

*Ben J. P. Salemans is a philologist with special interest in the use of the computer as a scientific tool in literary, philological and cultural-historical research. His Ph.D. project is a study of automated text-genealogical research. He works as a computer consultant at the UCI computer centre of the University of Nijmegen.*



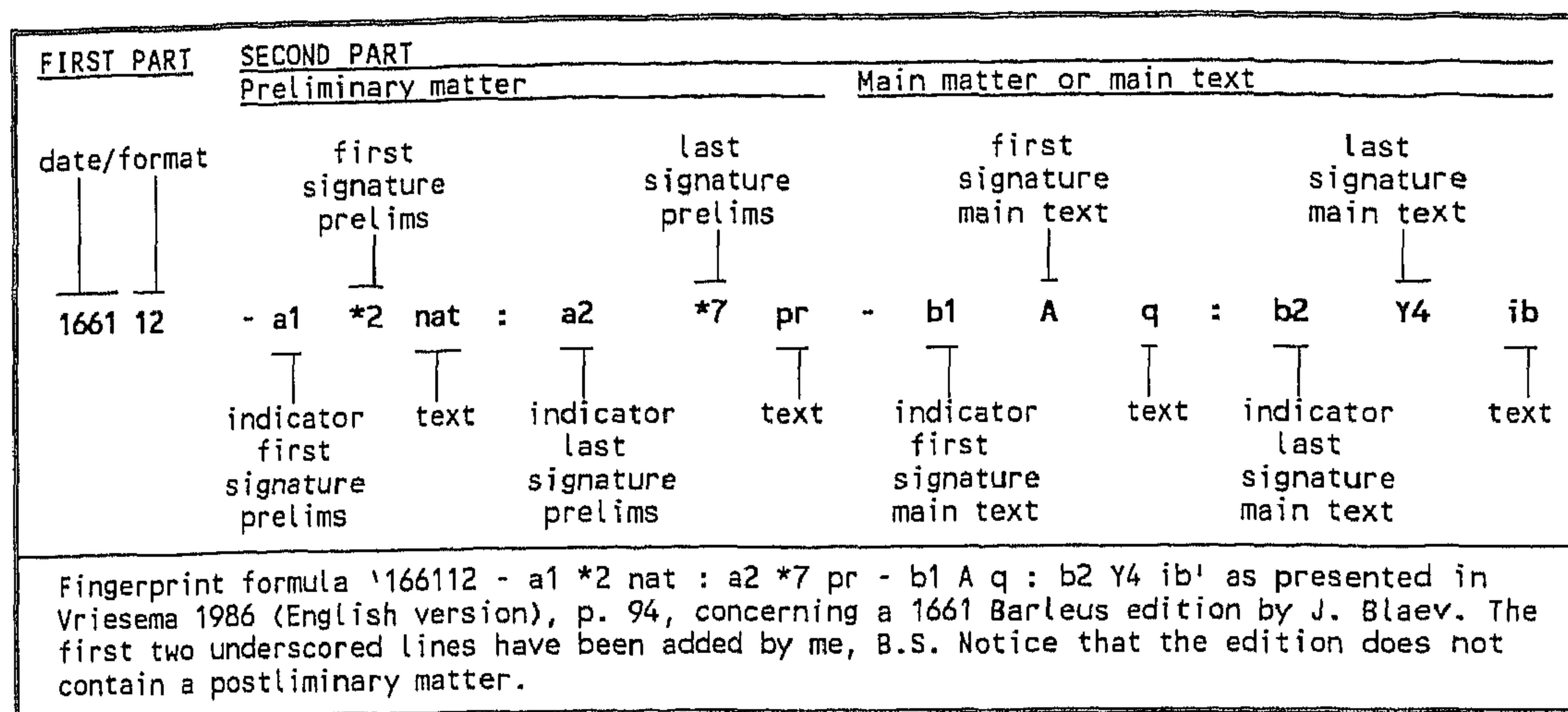


Fig. 1. Illustration of several components of a fingerprint formula.

octavo edition of the year 1560. The second part of the formula is more complex. It offers us information on the three parts an edition may show: (i) the preliminary matter, (ii) the main matter or main text, and (iii) the postliminary matter. Often, the preliminary and the postliminary matters do not occur in an edition. The signatures – characters at the bottom of certain pages which help bookbinders to place the gatherings (folded sheets) of a book in the right order – of the first and last gatherings are mentioned in the formula. To denote the position of a signature in an edition, each signature in the formula is preceded by a unique indicator. “a1” is the indicator for the first signature of the preliminary matters, “a2” the last signature of these prelims; “b1” is the indicator for the first signature of the main matter, etc. The page signatures are used by the fingerprint formula developers as fixed orientation points in the description of an imprint. Immediately after the signatures, the fingerprint includes the characters located directly above the signatures in the imprint. To find the relevant characters, imaginary vertical lines are drawn to the left and right of the signatures. The complete characters (generally, parts of characters are not allowed) between these lines, are included in the fingerprint formula.

In Vriesema (1986a) Figure 1 is used to illustrate the several components of a fingerprint formula: “166112 – a1 \*2 nat : a2 \*7 pr – b1 A q : b2 Y4 ib.” For example, the fingerprint “165204 – b1 A2 en\$b : b2 L uw” describes a quarto edition of the year 1652. It does not show preliminary or postliminary matters, since the

indicators “a1 ... a2” and “c1 ... c2” are not present in the formula. The first signature of the main text is “A2.” In the fingerprint it is preceded by the indicator of the first signature of the main text, “b1.” Directly above “A2” in the edition the characters “en\$b” can be found, in which the dollar sign represents one or more blanks. Sometimes another character may be used in a fingerprint to express a blank. The last signature of the main text is “L.” In the edition the characters “uw” stand just above it.

As Gerritsen explains in the introductory chapter of Schuytplot (1987, p. vii), there are good bibliographical reasons for these (STCN) fingerprint formulas. Fingerprints include elements of an edition, which have no textual meaning and have only a function for typesetters and bookbinders: the signatures. Even when a text is reprinted accurately, rule by rule, there is little chance that the non-textual signatures, which were placed after the text had been prepared, will emerge at the same place as in the previous edition. Moreover, a fingerprint formula concerns the signatures of several pages of an edition and the characters directly above them. Such formulas are intended to be as unique a tool for identification purposes as the human fingerprint.

### Limitations of STCN Fingerprint Formulas

Although the reasoning behind the fingerprint formula is solid, it is doubtful that the result will always be



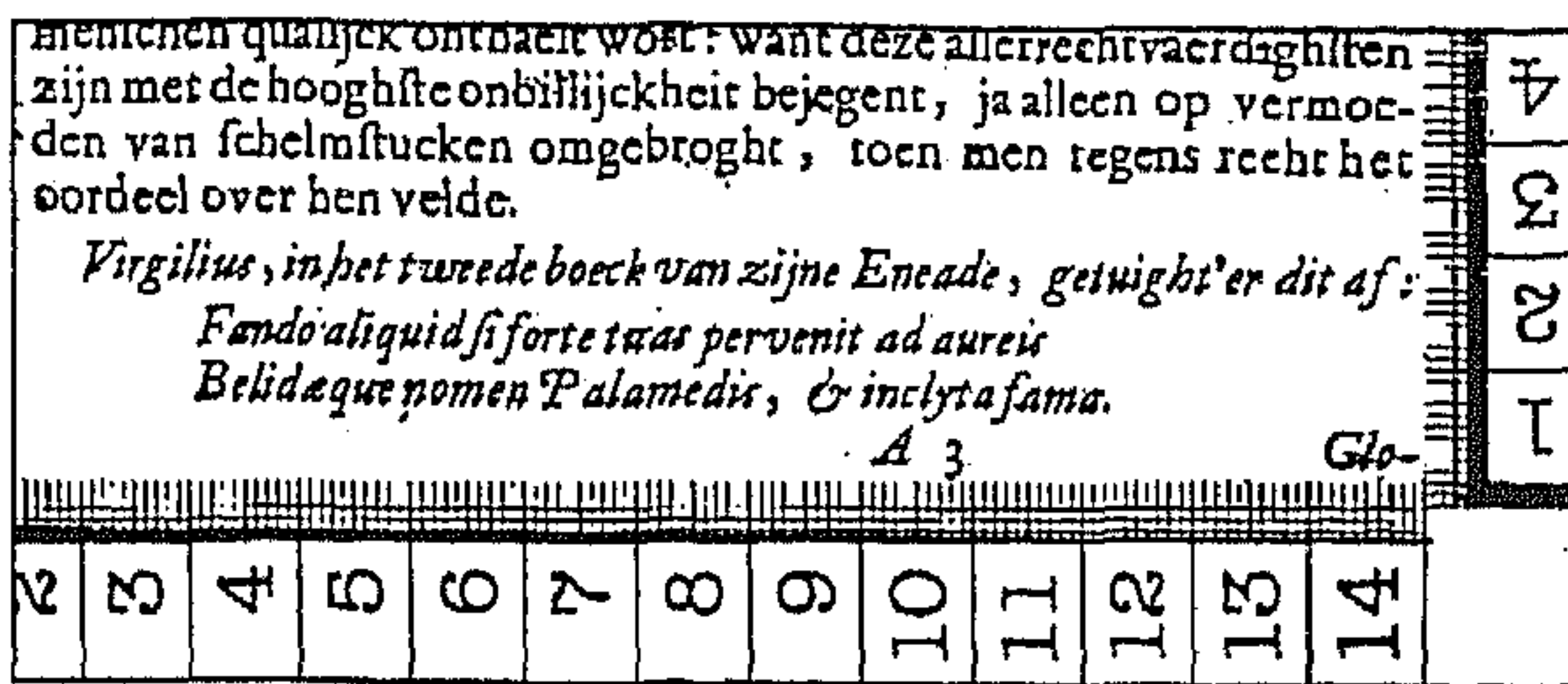


Fig. 2(a). "Duinkerken Palamedes," sign. A3 (= duina3.pcx); as in other Figures, a centimetre scale has been added at the borders.

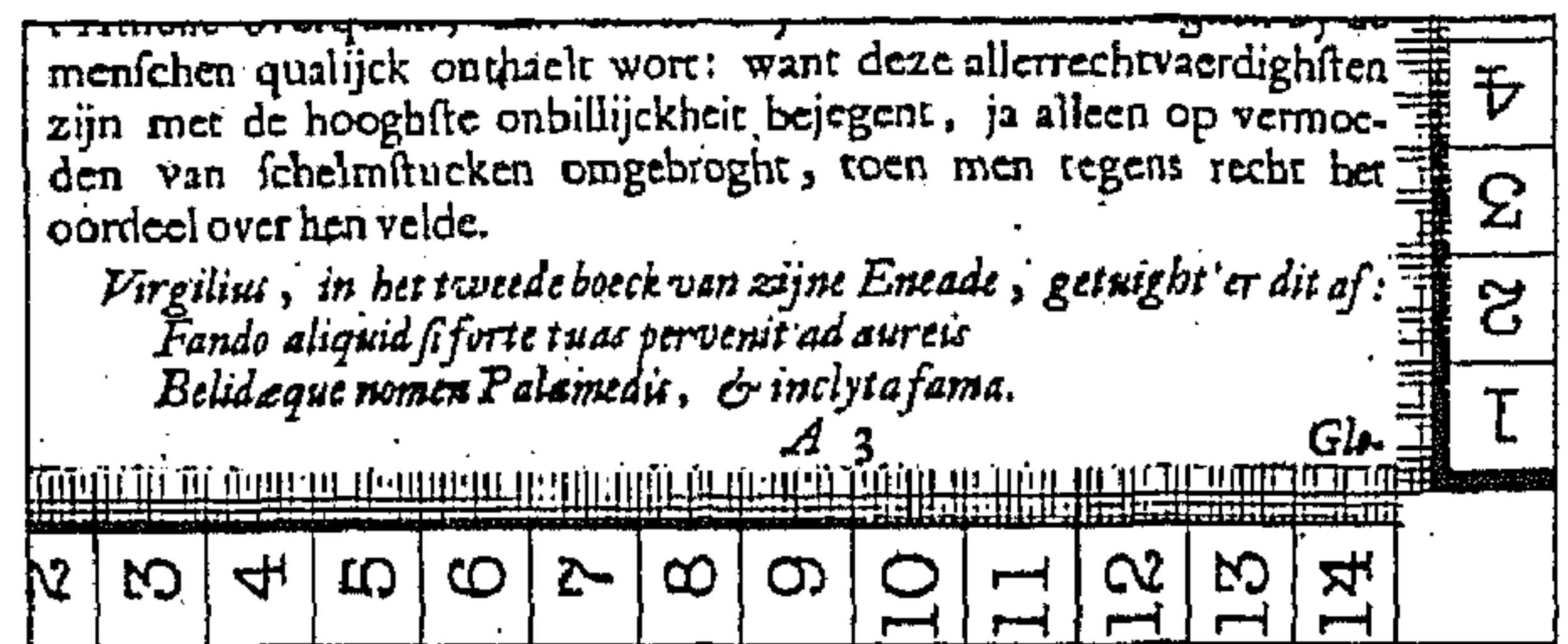


Fig. 2(b). "Cheap Palamedes," sign. A3 (= cheapa3.pcx).

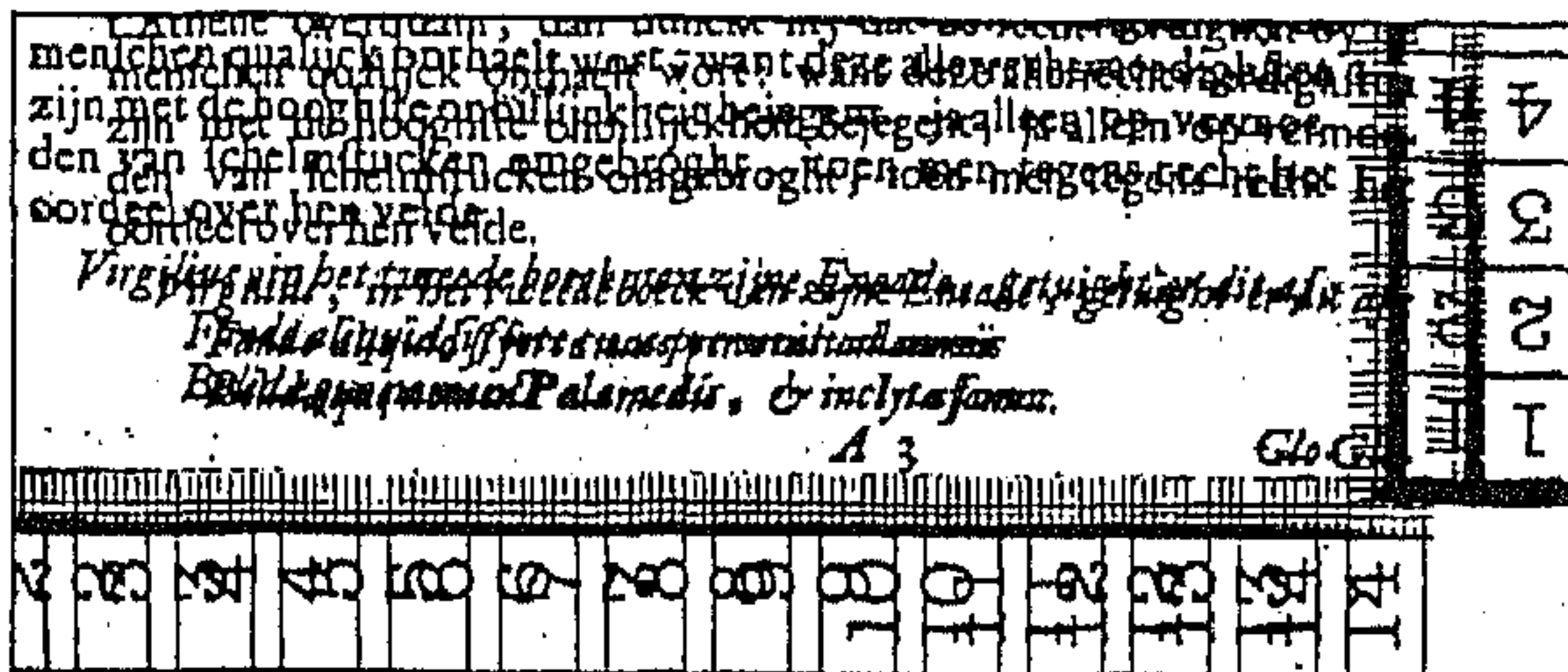


Fig. 2(c). Figures 2a and 2b combined (= comba3.pcx).

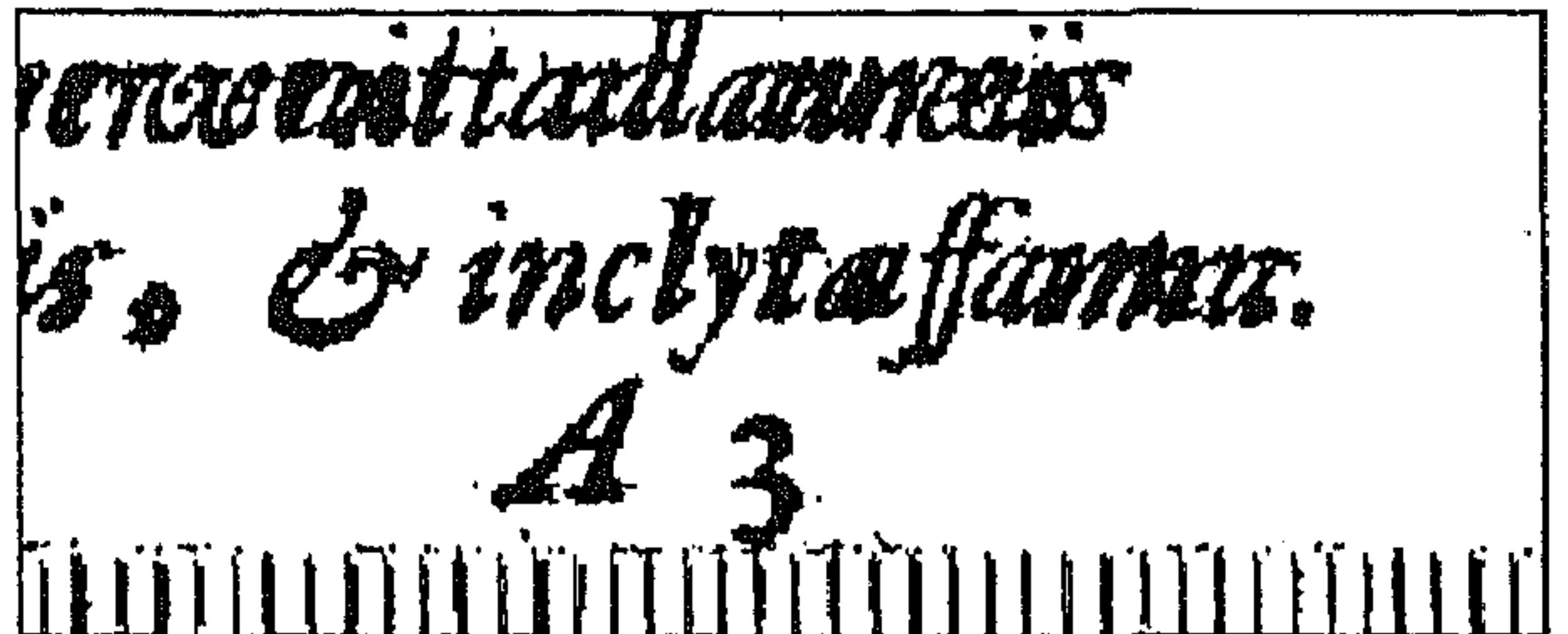


Fig. 2(d). Enlarged detail of Figure 2c.

an accurate description of a book edition. In point of fact, an STCN fingerprint formula sometimes offers a limited and potentially ambiguous description of an edition.

First, a fingerprint formula involves limited parts of certain pages in an edition. Only characters in the single line directly above the prescribed signatures are considered. The main question is whether to accept Gerritsen's contention that there is little chance of non-textual signatures emerging under exactly the same characters in different editions. I am convinced that his claim is not totally valid. This can be seen in Figures 2a and 2b, which show corresponding fragments of different editions of "Palamedes," a seventeenth century drama by Joost van den Vondel.

The characters above the signatures "A3" are equal in both editions: "ncly". Logically, it cannot be denied that two different editions might lead to the same fingerprint formula – a fact which brings into question its accuracy as a mechanism for identifying editions – even though it must be noted that a fingerprint formula includes several page fragments, or fingerprint areas, and that the chances of error are thus minimized. However, to detect the differences between the two editions more clearly, we can use the computer to place Figure 2b over Figure 2a, as if they were printed on transparencies. The "A3" signatures of both editions have to overlap. The result is pictured in the combined Figure 2c. There is no doubt that the two editions are different

in spite of the fact that they have similar fingerprint areas.

When we take advantage of the computer to enlarge a part of Figure 2c, as seen in Figure 2d, we can see more details. It is clear, for instance, that the characters "l" of "ncly" are different. The "l" of one edition is not as cursive as the "l" of the other edition.

Second, the rule mentioning only complete characters or signs above the signatures as constituents of a fingerprint is rather vague. In Vriesema (1986a, pp. 98 f.) this rule is formulated as follows:

4.1. The piece of text appearing above the prescribed signature is recorded, i.e. those characters that fall wholly or virtually wholly within the prescribed limits. ("Virtually wholly" means: where it is impossible to decide whether the character does or does not fall wholly within the limit.) Characters which only partly fall within the limit are disregarded. If a signature falls under only a part of one character or under parts of two successive characters, these parts count as whole characters (...).

In the Dutch Handleiding STCN (1988, p. 67) this rule has been changed into (my translation):

87. 1. The piece of text appearing above the prescribed signature is recorded, i.e. those characters that fall completely within the prescribed limits; characters which do not fall completely, with one hundred per cent certainty, within these limits are disregarded. However, if a signature falls under



only a part of one character or under parts of two successive characters, these parts count as whole characters (...).

The difference between the two rules is significant. If the bibliographer has doubts whether a character stands under a signature completely, the first rule states that it may be incorporated into the fingerprint, while this is forbidden by the second rule. The latter, more recent, rule seems to be more attractive, since the vague expressions “virtually wholly” and “impossible to decide” have been eliminated. However, it is still not accurate enough. Bibliographers will each have a personal interpretation of the meaning of “one hundred per cent certainty.” The result may be that two bibliographers produce different fingerprint formulas to describe the same text edition. Moreover, we must realize that a printer will ink the leaden types several times. Sometimes the characters in one printing of an edition may be much bolder than the characters in another printing of the same edition. The different boldness of the printed characters may result in different fingerprints for the same editions.

Furthermore, the last rule of both quoted passages leads to problems of interpretation. Suppose that one part of a fingerprint formula indicates that the characters “ma” are to be found above the first main text signature “A”: “b1 A ma”. Suppose, furthermore, that the string “ma” is part of the word “smart”. The question now arises as to how to interpret “ma” in the formula. For example, does it mean that parts of the “s” and “r” of “smart” also fall under the signature “A”, and are therefore not mentioned in the fingerprint? Or, on the other hand, does “ma” mean that the signature “A” falls under small parts of the successive characters “m” and “a”? In other words, the range or meaning of the text “ma” in the fingerprint is unclear and ambiguous.

### Fingerprint Formulas in Practice: Editions of Vondel’s “Palamedes” (1652)

It is possible to make other arguments against fingerprints, such as the occurrence of “parallel editions” (see Handleiding STCN, 1988, § 59.2). I hope, however, that I have sufficiently supported my thesis that STCN fingerprints are sometimes inaccurate and ambiguous.

Let us consider a practical example. Joost van den Vondel is a well-known seventeenth-century Dutch author. I have two versions of his drama “Palamedes” printed in 1652; the first edition dates from 1625. They

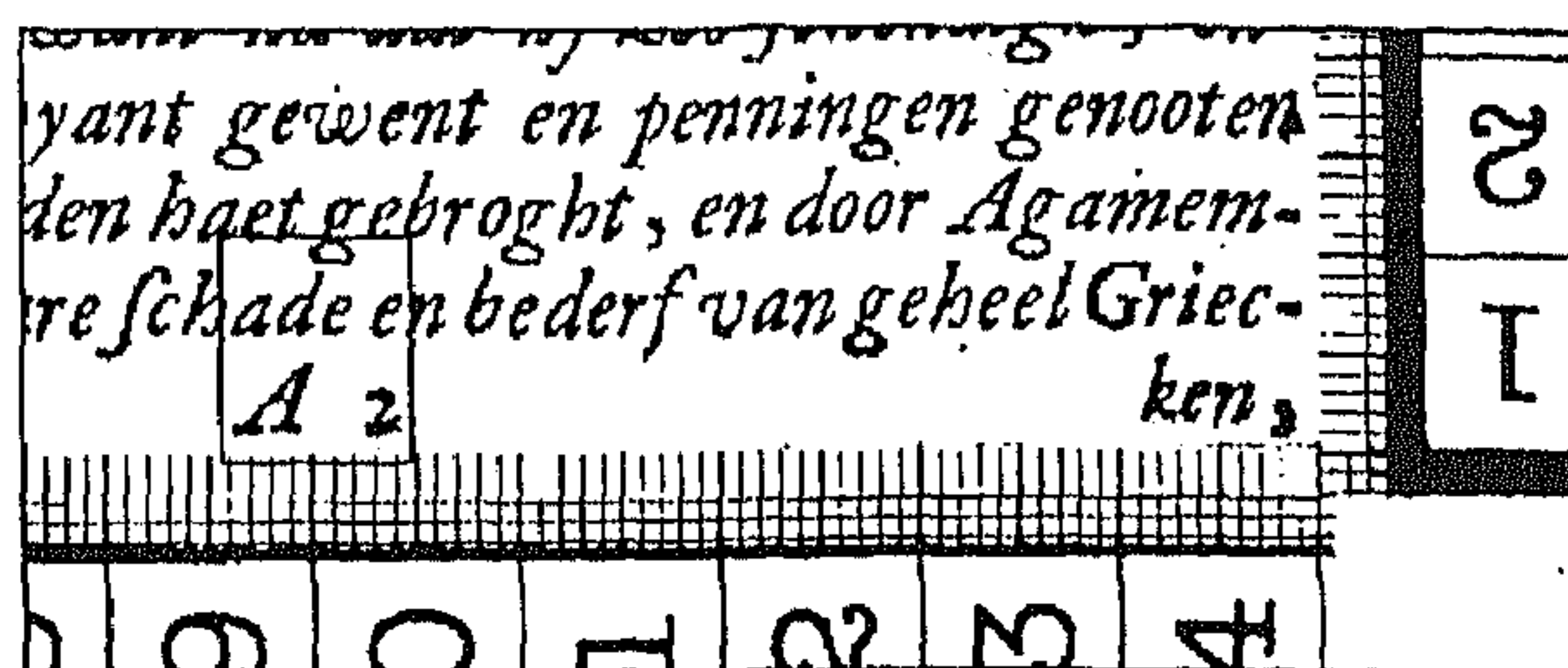


Fig. 3(a). “Duinkerken Palamedes,” sign. A2 (= enlarged detail of duina2.pcx); in Figures 3a, 3b, 4a and 4b, a rectangle is added to denote the fingerprint area.

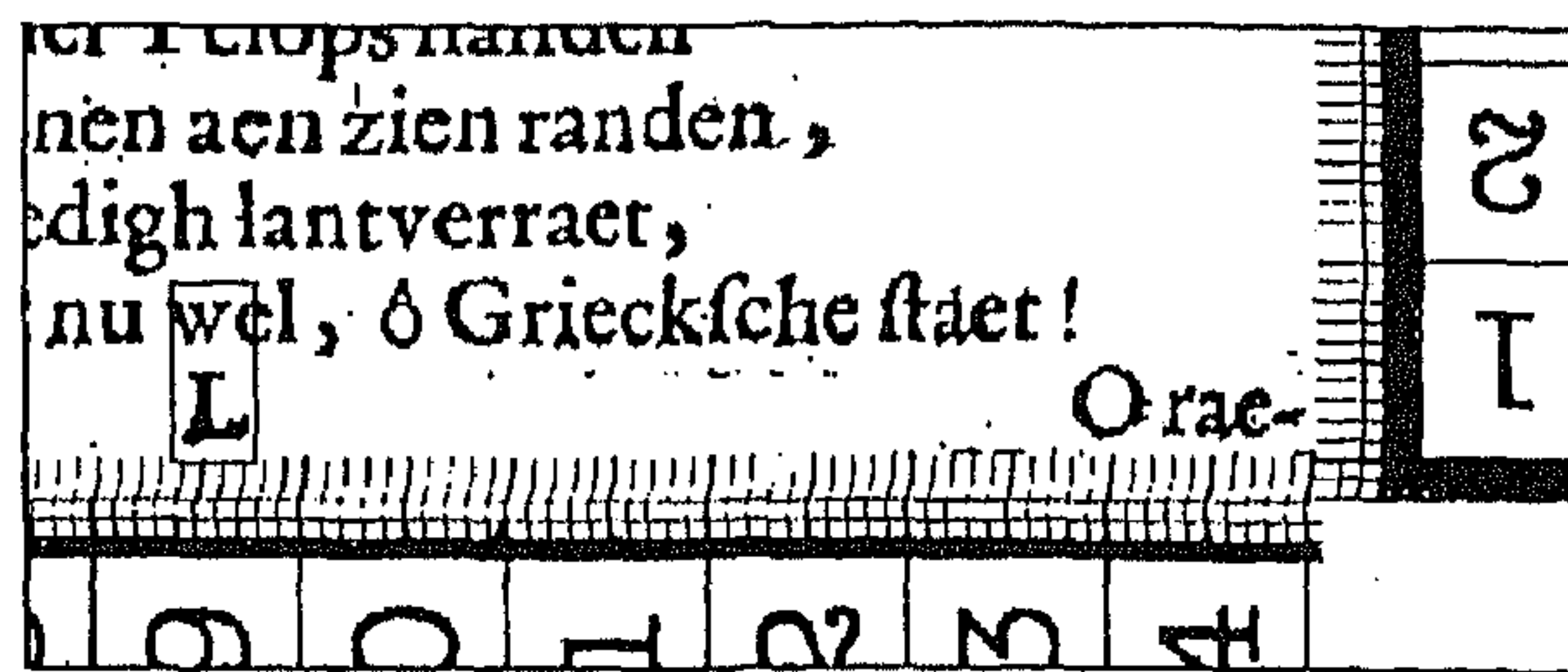


Fig 3(b). “Duinkerken Palamedes,” sign. L. (= enlarged detail of duinl.pcx). Figures 3a + 3b lead to the fingerprint: 165204 – b1 A2 ade\$e : b2 L we.

both occur in convolutes, which also contain other Vondel dramas. One convolute comes from the library of Anton van Duinkerken. I bought the other one for only a few guilders at an auction. I will refer to them here the “Duinkerken Palamedes” and the “Cheap Palamedes.”

According to the title pages both editions are printed “t”AMSTERDAM, Voor Abraham de Wees, op den Middeldam. 1652.” When we compare the title pages of both editions, we immediately see that they cannot have been printed during the same printing process. In the Duinkerken Palamedes, among others, blanks occur between the characters of “AMSTERDAM,” which are not present in the Cheap Palamedes.

In the Nijmegen and Amsterdam university libraries there are two catalogues of Vondel’s books, both published on the occasion of his 400<sup>th</sup> birthday: Arpots (1987) and Schuytvlot (1987) respectively. These catalogues offer, among others, fingerprint formulas of the descriptions of Vondel editions. Of course, fingerprints are not present in Unger’s standard Vondel bibliography (1888). It is remarkable that neither Arpots nor Schuytvlot give direct information on how fingerprint formulas should be interpreted. Most buyers of the catalogues – only a few will be bibliographers – will not understand the fingerprint formulas without an introduction. Both Arpots (personal information) and Schuytvlot have used Vriesema (1986b) as



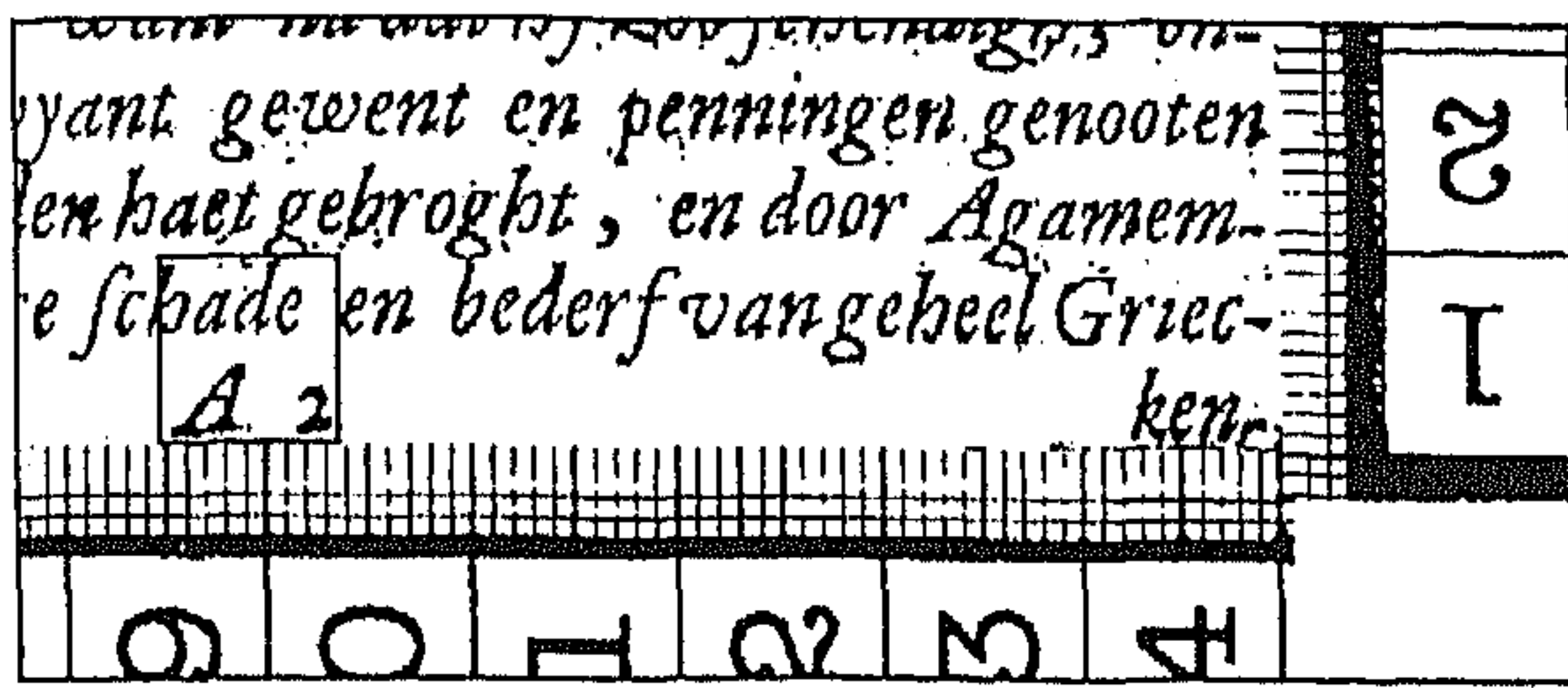


Fig. 4(a). "Cheap Palamedes," sign. A2 (= detail of cheapa2.pcx).

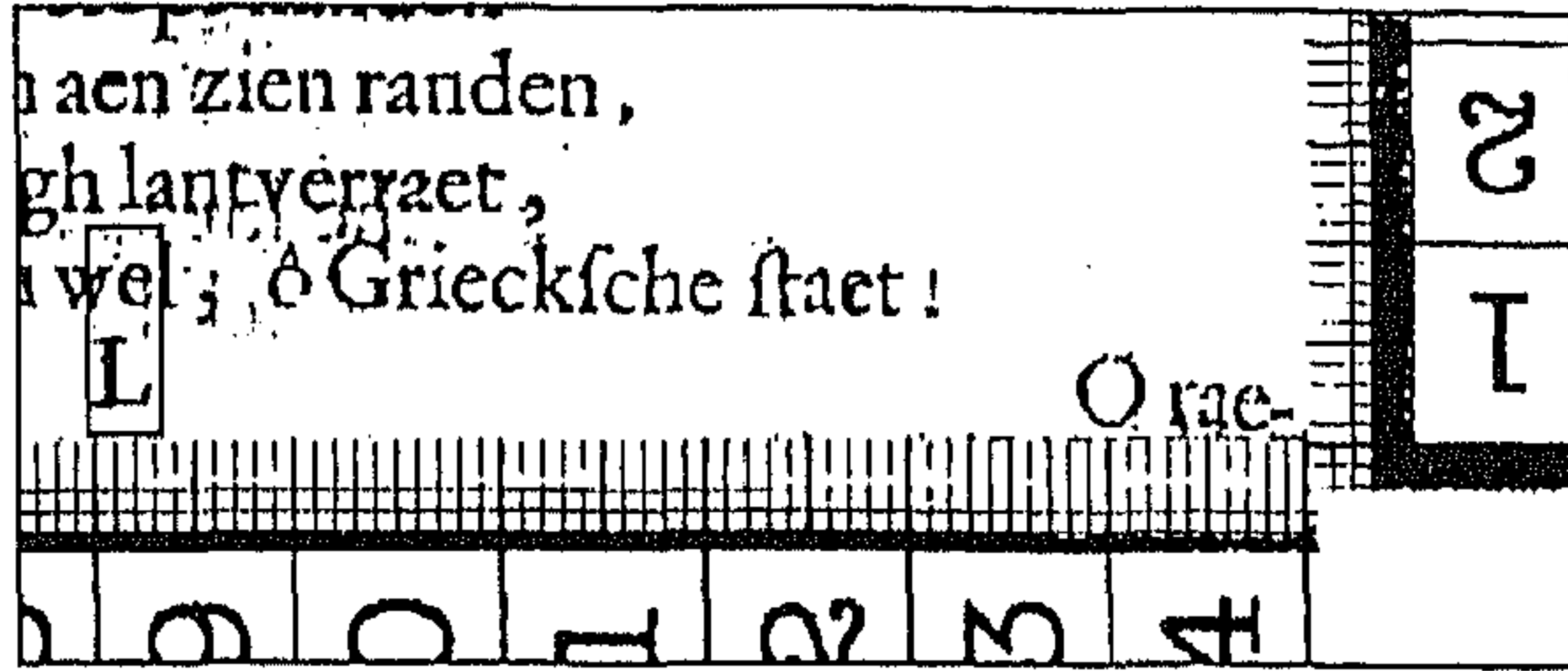


Fig. 4(b). "Cheap Palamedes," sign. L (= detail of cheapl.pcx); Figures 4a + 4b lead to the fingerprint: 165204 – b1 A2 ade\$ : b2 L e.

a guidance in developing fingerprints. If they had been able to use the rules of the Handleiding STCN (1988) their fingerprints would probably have looked quite different. Furthermore, in Vriesema (1986a,b) some examples of developing fingerprint formulas are presented, including the five Palamedes editions of 1652 present in the Royal Library in The Hague. Therefore, we have three fingerprint sources with which to compare our Palamedes editions.

We are able to develop our own fingerprint formula for the Duinkerken Palamedes. The first part of the formula is simple: "165204". No preliminary or postliminary matters occur. Logically, we can only use the first and last signatures of the main text, respectively "A2" and "L". The environments of the signatures are shown in Figures 3a and 3b.

Above the signature "A2" in Figure 3a we read the two words "schade en". To find out which characters have to be mentioned in the fingerprint we draw vertical lines – or, as I did in this and in some other figures, a rectangle – out of the left and right borders of the signature. The first part of the main text ("b1") above the first signature "A2" is "ade\$e". We should not be too impressed by the presented fingerprint area in the rectangle. For instance, if its right vertical line is moved only fractionally to the left, the "e" of "en" is possibly hit. Then the fingerprint part has to be "ade\$".

Unger 121	Vriesema 1	: 165204 - b1 A2 en\$b : b2 L uw
	Arpots 50	: 165204 - b1 A2 \$en\$b : b2 L uw
	Schuytvlot 447	: 165204 - b1 A2 en\$be : b2 L u\$w
Unger 122	Vriesema 2	: 165204 - b1 A2 de\$e : b2 L we
	Arpots 51	: 165204 - b1 A2 de\$ : b2 L we
	Schuytvlot 442	: 165204 - b1 A2 ade\$e : b2 L we
Unger 123	Vriesema 3	: 165204 - b1 A2 ade\$e : b2 L e
	Arpots 52	: 165204 - b1 A2 ade\$e : b2 L e
	Schuytvlot 443	: 165204 - b1 A2 ade\$e : b2 L e
Unger 124	Vriesema 4	: 165204 - b1 A2 ade : b2 L we
	Arpots 53	: 165204 - b1 A2 ade : b2 L we
	Schuytvlot 444	: 165204 - b1 A2 Ade\$ : b2 L e
(Sic! A capital A in the middle of a word? BS)		
Unger 125	Vriesema 5	: 165204 - b1 A2 ade\$ : b2 L \$w
	Arpots	: not mentioned; not available
	Schuytvlot 445	: 165204 - b1 A2 ade\$e : b2 L we
	Schuytvlot 446	: 165204 - b1 A2 ade\$ : b2 L \$w
<hr/>		
Duinkerken Palamedes		: 165204 - b1 A2 ade\$e : b2 L we
Cheap Palamedes		: 165204 - b1 A2 ade\$ : b2 L e

Fig. 5. Palamedes 1652 fingerprints of Vriesema, Arpots, Schuytvlot, compared with our Palamedes 1652 fingerprints.

Using the added rectangle in Figure 3b we see that the vertical lines hit parts of the "w" and the "e". When two signs stand partly above the signature, both signs have to be mentioned in the fingerprint formula. Therefore, the second part of the main text ("b2") above the final signature "L" is "we". This leads to the fingerprint of the Duinkerken Palamedes: "165204 – b1 A2 ade\$e : b2 L we". By analogy, with Figures 4a and 4b we develop the Cheap Palamedes fingerprint: "165204 – b1 A2 ade\$ : b2 L e".

### Comparing the Fingerprints with Vriesema, Arpots and Schuytvlot

As stated, Vriesema (1986a,b), Arpots (1987) and Schuytvlot (1987) offer us fingerprint formulas for several Palamedes 1652 editions. These formulas are presented in Figure 5. At the bottom of this figure the fingerprint formulas of the Duinkerken and Cheap Palamedes are given. Using Figure 5 we can draw several conclusions:

First, we observe that the fingerprints of the same editions, with the accompanying old Unger (1888) catalogue numbers, differ considerably. According to Van Selm (1988) the differences between the fingerprints of Arpots and Schuytvlot have apparently been caused by the fact that they simply do not know how to establish fingerprints. However, it is perhaps more likely that the fingerprint system leads to different fingerprints.

Second, we see that the fingerprint of the Duinkerken Palamedes matches not only with Schuytvlot 442, but also with Schuytvlot 445. Log-



ically, we conclude that it is likely that the University Library has the Duinkerken Palamedes edition. What is more important is the fact that Schuytvlot unknowingly – he does not pay any attention to it – proves that two different editions can in practise show the same fingerprint formula. Of course, I have to assume that Schuytvlot's fingerprint is correct, though this should be verified. At the moment, only a time-consuming visit to the Amsterdam university library could provide the answer. If we had a photograph of the relevant fingerprint area, the problem could be solved within seconds.

Third, if Schuytvlot's fingerprint no. 444 had not shown a capital "A" instead of a lower-case "a", it would agree with the fingerprint of the Cheap Palamedes. I fear that Schuytvlot's "A" has to be interpreted as an "a". Again, if we had a photograph of the page fragment concerned we could solve this capital A problem immediately.

It can be argued that a fingerprint offers only a partial bibliographical description of an edition. Fingerprints must be used in conjunction with other bibliographically descriptive elements to establish the (unique) identity of an imprint. Bibliographers could temper my criticisms of the fingerprint formulas by utilizing these other bibliographical elements. However, if I interpret Vriesema (1986a, p. 95) correctly, I fear that the comparison of STCN entries will be performed (automatically) with fingerprint formulas only:

*Matching*, i.e. bringing together copies appearing superficially to represent other different editions, or even different texts, can be most easily accomplished by getting the computer to gather all the identical fingerprints in a catalogue.

Besides, I hope that fingerprints which differ by only one or a small number of characters ("fuzzy matching"), will also be presented to the users of the STCN catalogue.

### First Proposal to Use Scanned Images of Texts to Identify Editions

It has been demonstrated that fingerprints sometimes turn out to be inadequate tools to describe and distinguish between different editions. The conclusion could be that we should improve them. We could choose, for instance, to allow parts of the signs above the prescribed signatures to be expressed in these formulas as

well. I fear, however, that this decision will result into more complex and user-unfriendly fingerprints.

A more obvious option is to maintain the current fingerprint formulas and add to them illustrations of at least the fingerprint areas, as shown in this paper. This would enable us to verify and interpret the presented formulas. The illustrations can be prepared with scanners. We know that scanners allow us to create computer photographs ("scans") of editions or drawings, etc. Figures 2a, 2b, 3a, 3b, 3c, 4a and 4b in this paper are examples of scans. Scans of fingerprint areas can easily be incorporated into a paper or electronic book catalogues and bibliographies. Furthermore, these scans can be distributed quite easily on floppies, cheap CD-ROMs or, preferably, via computer servers attached to the Internet. Members of the public can then manipulate and investigate the scans themselves on their own computers. For the manipulation of the scans in this paper I used the rather simple Windows software package *CorelDRAW* 3.0. The rectangles in Figures 3a, 3b, 4a and 4b, for example, were added to the scans with the help of this package. *CorelDRAW* also has a merging facility, which enables us to combine two scans into one new scan, as demonstrated in Figure 2c. This feature is very useful for the comparison of two or more fingerprint areas or other parts of text editions.

Once we have decided to use scanner technology in the process of edition identification, we should consider not strengthening the old, manual fingerprint method, but rethinking the identification problem from scratch.

Before we can use scans instead of, or next to, fingerprints, it must clear how an accurate analysis or comparison of computer scans should be performed. One fundamental question, for instance, is how differences between scans can be quantified accurately. At the moment it is not totally clear to me how such an analysis should be performed. Actually, this paper aims only to show that the comparison of scans is indeed possible. It is my hope that this will prompt computerized bibliographers to elaborate my proposal to use scans in the identification of text editions.

Evidently, many standards will first have to be established: the software and hardware to scan with, the scan type, the resolution, the exposure, the scaling to ensure that the scale of scans done in different places is the same, etc. Only with the adoption of standards will scans be able to serve, amongst other things, as comparable and objective computer descriptions of fingerprint areas. However, even after the standards



have been set, many practical problems will remain to be solved. Old books, for instance, are often bound so tightly, that it is difficult to make a good scan of a page fragment. This is clearly demonstrated by the wavy paper centimetre scale at the bottom of Figure 3b. One alternative is to make photographs of the pages concerned and then scan either the photos or the negatives. Another problem is how we should scan damaged or stained material. Not the least problem will be to convince the rare books librarians or collection owners that fragments of "their" books have to be scanned. Much work remains to be done.

### Acknowledgements

I would like to thank R. Arpots of the University Library of Nijmegen, E. Cockx-Indestege of the Royal Library in Brussels, J. Gilbert of the University Library of Maastricht and two anonymous reviewers of *CHum* for their comments on a previous version of this paper. With M. Bohnsack from IEM Inc. in Nijmegen I had several interesting conversations on some technical aspects of this paper. Together we successfully tested the CAD/CAM Intergraph software package Microstation, with its module IRAS-B. In the future this paper will possibly be translated into Dutch. The thoughts and ideas in this paper are strictly mine and not necessarily those of my employer, the University of Nijmegen.

### References

- Arpots, R. *Vondel in Nijmegen*. Nijmegen: Universiteitsbibliotheek, 1987.
- Foxon, David F. *Thoughts on the History and Future of Bibliographical Description*. Los Angeles: School of Library Service, University of California, 1970.
- Gaskell, P. A. *A New Introduction to Bibliography*. Corr. Ed. Oxford: Clarendon Press, 1979.
- Handleiding STCN. *Handleiding voor de medewerkers aan de STCN*. Tweede herziene uitgave. 's-Gravenhage, 1988.
- Karney, James. "Professional illustration." *PC Magazine*, 12, 6 (March 30, 1993), 275-307.
- Laufer, R. *Introduction à la textologie*. Paris: Larousse, 1972.
- Salemans, B. J. P. "Jacob Bathen, Printer, Publisher and Bookseller in Louvain, Maastricht and Düsseldorf, c. 1545 to c. 1557." *Quaerendo*, 19 (1989), 3-47.
- Schuytvlot, A. C. *Catalogus van werken van en over Vondel*. Bibliotheca Bibliographica Neerlandica. Vol. 25. Nieuwkoop: De Graaf, 1987.
- Selm, B. van. "De raadsels van de Vondeldrukken." *Dokumentaal*, 17 (1988), 19-25.
- Unger, J. H. W. *Bibliographie van Vondels werken*. Amsterdam: Muller, 1888.
- Vriesema, P. "The STCN Fingerprint." *Studies in Bibliography*, 39 (1986a), 93-100.
- Vriesema, P. "De STCN-Vingerafdruk." *Dokumentaal*, 15 (1986b), 55-61. (A translation of Vriesema 1986a into Dutch - BS).

### Note

- <sup>1</sup> Dedicated to the memory of Marcel te Wilt († 1993).