

The Disputed Authorship of a Medieval Text: A New Solution to the Attribution Problem in the Case of *Guillaume D'Angleterre*

MARIA SLAUTINA

Abstract: This paper presents a case of disputed authorship concerning a literary medieval text, the Old French romance *Guillaume d'Angleterre*, and describes a new attempt to use the mathematical method of authorship attribution called the "method of pattern recognition". After presenting an overview of previous approaches to the problem of the authorship of this text, I argue for the advantages of a statistical syntactic based method for authorship attribution of medieval texts. The method of pattern recognition consists in the identification of a domain of syntactic parameters and a measurement of the proximity or distance of texts as located in a multi-dimensional syntactic space. I find that the medieval text most likely belongs to Chrétien de Troyes, one of the most famous French authors of the twelfth century. I present for the first time an attempt to apply the method of pattern recognition to determine the authorship of a medieval text written in Old French.

Contributor: Maria Slautina completed her PhD in the joint fields of mediaeval French literature and mathematical linguistics, attached to the LASLAR laboratory of the University of Caen Lower Normandy and the Department of Mathematical Linguistics of the State University of Saint Petersburg. Her research had to do with the problem of attribution in mediaeval texts and their solution as suggested by techniques drawn from mathematical linguistics. Her current interests focus on the figure of the author in French medieval manuscripts, on the relationship between text and illumination, and on the emerging status of the author in medieval and early Renaissance literature.

1. Introduction.

The question of the authorship of *Guillaume d'Angleterre*, a French romance commonly dated to around 1170, has been discussed for more than a century by eminent medievalists in France and abroad. The main reason for this interest is the potential attribution of the text to Chrétien de Troyes, one of the most important French medieval writers and the so-called "father of French romance". Researchers are not only divided into two camps approving or disapproving Chrétien's authorship, they also entertain diametrically opposite opinions concerning the quality of the writing, in general reflecting their attitude towards its paternity. Some see in this hybrid of hagiography with the adventure romance another example of Chrétien's versatile genius; others think that the Champagne writer would never conceive such a parody of his Arthurian knight romances.

I would argue that there are objective criteria one can apply to a medieval texts to establish its provenance. I will describe such criteria and will demonstrate in this paper how a method of pattern recognition issuing from cybernetics will work for the attribution of literary texts. The method of pattern recognition consists in the identification of a domain of parameters and a measurement of the proximity or

Slautina, Maria. "The Disputed Authorship of a Medieval Text: A New Solution to the Attribution Problem in the Case of *Guillaume D'Angleterre*." *Authorship* 2.1.

DOI: <http://dx.doi.org/10.21825/aj.v2i1.760>

Copyright Maria Slautina. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

distance of texts as located in a multi-dimensional space. The method is holistic rather than analytic. Pattern recognition models make up part of the theory of pattern recognition and are normally included in the research domain of artificial intelligence. These models are used in many areas of analysis where there is need to classify different objects, phenomena, processes, signals, events, and so forth. Among many of their practical applications encountered in everyday life are included for instance face-recognising software in photograph applications and the classification of spam and non-spam e-mails by e-mail box filters. In the present case the method of pattern recognition is adapted to the classification of literary texts by creating models of them using specific parameters based on the syntactic structure of the text, when one extracts an extensive number of syntactic parameters from the compared texts and applies statistical criteria, such as the *t criterion of Student* and the measurement of Euclidian distances between objects-texts, in order to evaluate their proximity, once the texts are put in an X-dimensional space, where X is a number of relevant distinguishing parameters (see Graph 1 at the end of this article).

The method of syntactic pattern recognition has been successfully tested on different Russian and French texts: it has shown, for instance, that Emile Ajar and Romain Gary were the same author¹; resolved the famous attribution case of *And Quiet Flows the Don*, signed by the Nobel winner Mikhail Sholokhov, attributing most of it to his collaborator and “teacher” Mikhail Serafimovich; and proved or rejected many cases of attribution of anonymous or pseudonymous writings to Pushkin, Dostoyevsky, Mayakovsky and others.² Research into the texts signed by Molière showed that a large part of his work may belong to the pen of Corneille.³

1. Discussions surrounding the authorship of *Guillaume d'Angleterre*.

The idea of attributing the tale to Chrétien de Troyes stems from of the occurrence of the name “Crestiens” in the prologue. “Crestiens se veut entremetre, sans nient oster et sans nient metre, de conter un conte par rime [...]” (v. 1-3) begins *Guillaume d'Angleterre*. Verse 1 of the romance thus raises the problematics of our research: this “Crestiens”—is he Chrétien de Troyes, the well-known twelfth-century author who composed five still-popular romances?

Previous analysis of *Guillaume d'Angleterre*'s authorship has mainly occurred in a discussion format. The first big debate took place between Maurice Wilmotte and Frédéric-Joseph Tanquerey in the early twentieth century. Wilmotte⁴ is without doubt one of the most ardent defenders of the attribution of the romance to Chrétien de

¹ Valentina Chepiga, “Méthodologies croisées pour l’attribution des textes: la place de la génétique”. *Modèles linguistiques*. Tome XXX. Vol. 59. Toulon: Éditions des Dauphins, 2009. 101-132.

² Mikhail A. Marusenko, Boris L. Bessonov, Larisa M. Bogdanova, *В поисках потерянного автора (этюды атрибуции)* [In Search of the Lost Author (attribution etudes)], Saint-Petersburg, 2001.

³ Mikhail A. Marusenko, Elena S. Rodionova, “Mathematical Methods for Attributing Literary Works when Solving the “Corneille-Molière” Problem”, *Journal of Quantitative Linguistics*, 17:1, 2010, p. 30-54

⁴ Maurice Wilmotte, “Le conte de Guillaume d'Angleterre”, in *Le Moyen Age*, 2^{ième} année, 1889, p. 188-191 and “Chrétien de Troyes et le conte de *Guillaume d'Angleterre*” in *Romania*, t. 47, n^o 181-184, 1920, p. 1-38.

Troyes. He defends his thesis peremptorily, entering into the controversy along with another supporter of the same theory, Wendelin Foerster.⁵ The two researchers, while convinced of the correctness of the attribution, have differing views on the place of *Guillaume* in Chrétien's works. According to Foerster the romance belongs to the final period of his career, and reflects the decline of his creative forces. Wilmotte rejects the proposal in a categorical manner. He insists that the novel is an integral part of the work of Chrétien de Troyes: "The style of *Guillaume d'Angleterre* is no less rich or less diverse than other works of Chrétien".⁶ In his article from 1920 he presents *Guillaume* as an adventure novel and an aristocratic legend, and insists that it has all the typical features of the Champagne master's works. Wilmotte reveals in the text of *Guillaume* figures of speech occurring in Chrétien's romances: anadiploseos, chiasmus, synonyms, and pleonasm, as well as "rare" and "unconscious" rhymes he believes every author uses together with the "conscious" and conventional ones. Having selected 500 rhymes at random in *Guillaume*, Wilmotte demonstrated that about 50 of them were used by Chrétien in his works of incontestable authorship.

All these described findings provoked great scepticism on the part of Tanquerey.⁷ He agrees that *Guillaume* might be an adventure romance, but one of a completely different nature from the adventure romances signed by Chrétien de Troyes. First and foremost, Tanquerey bases his objections on the absence in the romance of the characteristic features of Chrétien de Troyes' style: subtle psychology of the characters, chivalry, and the specific use of the wonderful. While Chrétien uses monologues to show us the inner struggle of the hero before taking a decision or to discover his state of mind, and while those monologues are always highly emotional, the author of *Guillaume* uses this literary device as a decorative surface which does not clarify the inner life and thoughts of his characters. Nor is *Guillaume d'Angleterre* "un poème chevaleresque", of the sort that Chrétien is supposed to compose. We do not see Yvain, nor king Arthur and his court, nor tournaments, nor chivalrous exploits. We do not find any manifestations of beauty: no beautiful ladies and knights, no clothes and no lavish precious weapons, no dazzling celebrations. The critic considers significant this lack of the stylistic details so important to Chrétien. Finally, the nature of the wonderful in *Guillaume* is different from its character in the works of that writer. In the latter, it is a magic wonderful, while in *Guillaume* it is supernatural, such as the heavenly voice which gives the orders to the characters, and one would never meet such a pattern in the courtly romances. Tanquerey comes to the conclusion that, in terms of the analysis of these three essential themes, *Guillaume* is based on a completely different model. He admits, of course, that the poet could have written a work so very different from the others, and he also admits that his arguments are not sufficiently definitive to deny Chrétien's authorship, but he underlines that it is not possible to refer to the nature and to the genre of the romance to assign it to the Champagne poet. To respond to Wilmotte's assertion about the

⁵ Wendelin Foerster, *Sämtliche erhaltene werke*, M. Niemeyer, Halle, t. IV, 1899, p. 255-425.

⁶ Wilmotte, 1889, p. 190.

⁷ Frédéric-Joseph Tanquerey, "Chrétien de Troyes est-il l'auteur de *Guillaume d'Angleterre*?", in *Romania*, t. 57, n° 223-225, 1931, p. 75-116.

resemblance of versification processes in both corpuses of texts, Tanqueray undertakes a new study of rhymes, after which he concludes that if we take 500 verses in *Guillaume* and compare their rhymes with those of any other work of the same period (provided it is long enough; he compares it to *Amadas* and the Continuation of *Perceval*), there will always be a large number of rhymes called “rare”. For example, among the forty one rhymes chosen by Wilmotte, he discovered twenty-nine which were either in *Amadas* or in the *Continuation of Perceval*, or in both.

After this “battle of the giants” of early twentieth century medieval studies, a number of medievalists and critics took one or the other side, supporting or rejecting the previous arguments and basing the analysis on new considerations. Gustave Cohen supports the attribution, acknowledging the absence of Celtic motifs and courtly love and therefore placing the romance in the beginning of the literary career of the writer, but this time without Foerster’s pejorative context.⁸ For Charles Foulon *Guillaume* is an “aristocratic work par excellence” that can be considered as a hymn to the courtly society and its feudal organisation and therefore absolutely conforming to the philosophy of Chrétien, who during his lifetime had several patrons and protectors and consequently one goal—to please those rich and noble people constituting his audience.⁹ Fausto Rebuffat points out that Chrétien follows the plot of a narrative he had not invented, and we have to deal in this case, not with his own creation, but with a transformation of a preexisting story.¹⁰ The writer has to follow the pre-existing narrative and cannot insert Arthur and his knight into the plot, but he does the revision in his personal style, “embroidering” by his hand the established pattern. Howard Robertson joins Rebuffat’s arguments although he admits that the “lack of conclusive evidence forces the decision that the poem is anonymous”.¹¹ Nevertheless he enters into polemic with Tanqueray and finds that the “merveilleux magique” of Chrétien de Troyes’ romances corresponds quite well to the wonderful as found in the lives of saints.

Jean Frappier, on the other hand, is skeptical about Chrétien de Troyes’ authorship and considers the art of the author of *Guillaume* in creating the plot far below the Champagne writer’s talent, although admitting the author’s undeniable literary capacities.¹² The most important argument against the attribution is Chrétien’s own enumeration of his works in *Cligès*, as Frappier can’t agree that a writer of his level would compose such a narrative after *Cligès* and in the middle of his Arthurian cycle. Alexandre Micha assents to the inferior quality of *Guillaume* and finds that it doesn’t

⁸ Gustave Cohen, *Chrétien de Troyes et son œuvre*, L. Rodstein, Paris, 1948.

⁹ Charles Foulon, “Les tendances aristocratiques dans *Le Roman de Guillaume d’Angleterre*”, in *Romania*, t. 71, 1950, p. 222-237.

¹⁰ Fausto Rebuffat, “La posizione e il valore del *Guillaume d’Angleterre* nella produzione poetica di Chrétien de Troyes”, in *Annali della facoltà di filosofia e lettere dell’Università statale di Milano*, vol. IV, fac. 1, gennaio-aprile, 1951, p. 23-34. Two possible sources are usually mentioned: *the story of Apollonius of Tyre* and *the Life of St. Eustache*.

¹¹ Howard R. Robertson, “The Authorship of the *Guillaume d’Angleterre*”, in *Romance Notes*, volume IV, 1962-63, p. 156-160, quotation page 157.

¹² Jean Frappier, *Chrétien de Troyes, l’homme et l’œuvre*, Paris, Hatier Boivin, 1957, rééd. 1968.

have the noble inspiration and the good taste of the authentic works by Chrétien.¹³ But Paul R. Lonigan supports the attribution and proposes to consider the works of Chrétien as the production of a “literary man” from whom we may see “love of parody, artistic rivalry, innovation and spoof”.¹⁴ He finds common tricks in his Arthurian romances and in *Guillaume*: scenes of non-recognition, concealment of information from the reader, “fortuitous coups”, and the symbolic use of objects.¹⁵ Finally, Norris J. Lacy doesn’t believe we would ever find a definitive answer to the question of this attribution but isn’t ready to deny *a priori* the authorship to Chrétien, as the author of *Guillaume* was a writer of a certain stature and the romance does correspond to the narrative construction of the Champaign author—thesis, antithesis, synthesis.¹⁶ Ruth Harwood Cline, admitting that it might well be the same Chrétien, sees “remarkable coincidences” between English toponyms described in *Cligès* and in *Guillaume d’Angleterre*, that can be explained by the kinship between Chrétien’s Champaign protector Count Henri and his half-brother Abbot Hugh of Lagny, who had connections with Norfolk, Bury St. Edmund, Windsor and Lagny.¹⁷

One cluster of researchers decided to turn away from the analysis of literary motifs, and to base their authorship reviews rather on the linguistic analysis of the stylistic features of the text. Arnulf Stefenelli thinks that the lexical core and the frequency of forms in play when there is a possibility of lexical choice may help reveal distinctive style marks.¹⁸ He concludes that if authorship can be based on the language specificities of Chrétien de Troyes, *Guillaume* will be his authentic romance, as unlike his contemporaries he has always chosen the most modern form of words: to “*mollier*” he prefers “*fème*”, to “*trametre*,” “*enveier*”, to “*cite*,” “*ville*”. He also likes to use pairs of synonyms like “*remembre*” and “*sovenir*”, which is an unusual process that appears only once among contemporary works, in the *Vie de Saint Thomas Becket* by Guernes de Pont-Sainte-Maxence. Stefenelli sees “indisputable” concordances between the languages of the Champagne writer and of the author of *Guillaume* in the matter of choice of lexemes and synonyms, leading him to conclude that these authors should be the same person. He reveals some discordancies too; for example, in *Guillaume* the verb “*gracier*” is used, while in his romances Chrétien prefers the word “*remercier*,” although this choice may be due to nuances of meaning (“*gracier*” means above all “*remercier Dieu*”). Maurice Delbouille presents a work based on the study of “banal” rhymes, going against previous

¹³ Alexandre Micha, *Le roman jusqu’à la fin du XIIIe siècle, Tome I (partie historique)*, sous la direction de Jean Frappier et de Reinhold R. Grimm, Heidelberg, 1978.

¹⁴ Paul R. Lonigan, “The Authorship of the *Guillaume d’Angleterre*: a New Approach”, in *Studi Francesi*, maggio-dicembre, 47-48, 1972, p. 308-314, quotation page 309.

¹⁵ The same kind of analysis and conclusions are presented in the article of Harry F. Williams, “The Authorship of *Guillaume d’Angleterre*”, in *South Atlantic Review*, t. 51, 1987, p. 17-24.

¹⁶ Norris J. Lacy, *The Craft of Chrétien de Troyes: An Essay on Narrative Art*, appendix “The composition of the *Guillaume d’Angleterre*”, Leiden, 1980, p. 118-124.

¹⁷ Chrétien de Troyes, *Cligès*, translated by Ruth Harwood Cline, University of Georgia Press, 2000, 217 p., p. 13.

¹⁸ Arnulf Stefenelli, “Die Autofrage des *Guillaume d’Angleterre* in lexikalische Sicht”, in *Verba et Vocabula: Festschrift E. Gamillscheg*, Munich, 1968, p. 579-591.

attempts to highlight the “original” ones.¹⁹ A rhyme is considered banal if it is repeated at least once in the text. Delbouille selects rhymes of this type in three romances of Chrétien de Troyes, in two romances of Gautier d’Arras, in the *Roman de Troie* by Benoît de Sainte-Maure, and in the works of Marie de France, and then compares the rhymes obtained from these works with those of the same type in *Guillaume d’Angleterre*. Three groups emerge: in the first are the romances of Chrétien and *Guillaume*; in the second are the works of Gautier; and in the third are the *Roman de Troie* and the poems of Marie de France.

Linguistic techniques of analysis gave rise to a number of studies claiming to refute Chrétien’s authorship too. Wolfgang Brand denies the attribution of *Guillaume* to Chrétien de Troyes, basing his conclusion on the fact that the two favourite techniques of the writer - entrelacement (*Reihentechnik*) and duplication (*Doppeltechnik*) - are absent in the text, while the romance is long enough to deploy at least one of these devices.²⁰ Brand also has objections to the lexical method of Stefenelli as it is based on a very limited number of words, and suggests that a similar study should be applied to the entire lexicon used by contemporaries of the writer. Lars Lindvall uses *Guillaume d’Angleterre* to compare its stylistic and syntactic characteristics with those of Chrétien’s works.²¹ When Lindvall compared all six romances (five works by Chrétien de Troyes and *Guillaume d’Angleterre*), he found that two of them have absolutely nothing in common with the others in terms of selected syntactic markers. *Erec et Enide* occupies the farthest point from the main corpus and *Guillaume* is situated between *Erec* and the other romances. Lindvall’s method does not support the idea of Chrétien de Troyes’ attribution; in this case, however, it would be logical to deny the attribution of *Erec*, the only text in which the poet explicitly mentions his full name. The opinion of Domenico d’Alessandro is based on the study of descriptions: he makes the link between the complexity and the structure of the text and the number of descriptions, concluding that the structure of *Guillaume* is much less complex than that of the other romances.²² For example, in *Guillaume* d’Alessandro finds only one example of the explicit introduction of a description; the rest of them are inserted *ex abrupto*. He examines the types of description and described objects and equally comes to the conclusion that there are not many similarities between the works. François Zufferey calls our attention to Picardian features of the text that he considers to be crucial in the attribution of *Guillaume*,

¹⁹ Maurice Delbouille, “À propos des rimes familières à Chrétien de Troyes et à Gauthier d’Arras”, in *Mélanges Lecoy*, Paris, 1973, p. 58-65.

²⁰ Wolfgang Brand, *Chrétien de Troyes : Zur Dichtungstechnik seiner Romane*, Wilhelm Fink Verlag, München, 1972, p. 202-212.

²¹ Lars Lindvall, “Structures syntaxiques et structures stylistiques dans l’œuvre de Chrétien de Troyes”, in *Romania*, t. 102, n° 4 (408), 1981, p. 456-500.

²² Domenico d’Alessandro, “Guillaume d’Angleterre et Chrétien de Troyes : un’analisi comparata del descrittivo”, in *Annali dell’Istituto Universitario Orientale di Napoli, Sezione Romanza*, t. 29, 1987, p. 349-56.

because it leads to the conclusion that the author wasn't from Champagne but from Picardy.²³

Without completely denying the possibility of Chrétien de Troyes being the author of *Guillaume*, Anthony J. Holden, the editor of manuscript C, is very reserved in this regard, stating that “the problem is not resolved, and we believe it will never be a definitive way”, and that “the identity of our author with Chrétien is far from being assured, [...] and the arguments against the attribution to Chrétien are at least as strong as those in his favour”.²⁴ Nevertheless, in the preface to his edition of the romance, Holden speaks of the attribution and makes comparisons. He proceeds to a comparative study of the subordinate clauses in *Guillaume* and in *Lancelot* that gives divergent results: on the one hand, it is possible to see “remarkable” agreements between the two texts which are in “a single state of language”, while on the other hand, “against the background of common usage stand a number of striking contrasts of specific uses”. The frequency of the different types of subordinate clause is relatively the same: twice as many in the 6149 verses of *Lancelot* (without the portion written by Godefroi de Leigni), as compared to the 3306 verses of *Guillaume*. However, there are differences: the hypothetical clause introduced by “mais que” is twice less frequent in *Lancelot*, while the consecutive negative is equally much more common in *Guillaume*. The result is the same from the analysis of conjunctions: if there are many similarities, there are also big differences (for instance, the conjunction “que que”, very characteristic of the language of Chrétien according to Holden, doesn't occur in *Guillaume* at all). The same differences concern the versification and the vocabulary. Holden concludes that the language of the author of *Guillaume* is not very different from the language of Chrétien; however, it concerns primarily similarities of vocabulary, which is more prone to imitation and to conformity. If we proceed to the analysis at the level of sentential rhythm, the results are different and the divergences are more remarkable.

The latest edition of the text of the romance based on the manuscript *P* was published in 2007 by Christine Ferlampin-Acher.²⁵ The name of Chrétien de Troyes with a question mark in brackets might make us think that the editor approves the attribution, but this is not the case. She supports the opinion that without the name in the prologue, serving as a basis for the authorship, such an idea would never have arisen among scholars. Besides, *Guillaume* shouldn't be dated at the end of the twelfth century, but is “definitely a posterior romance, entering [...] the margins of the Arthurian world and demonstrating a renewal of the genre that would belong perhaps more to the early thirteenth century rather than the twelfth.”²⁶

It is useful to make this brief survey of the previous scholarship in the field of *Guillaume* authorship, insofar as we can witness a progressive development of the tools

²³ François Zufferey, “La pomme ou la plume : un argument de poids pour l'attribution de *Guillaume d'Angleterre*”, in *Revue de linguistique romane*, t. 72, n° 285-286, janvier-juin 2008, p. 157-208. One of the manuscript presents indeed Picard features but it can be explained by an intervention of a Picard copyist.

²⁴ Chrétien, *Guillaume d'Angleterre*, édition critique par A. J. Holden, Genève, 1988, quotation p. 16.

²⁵ Chrétien de Troyes (?), *Guillaume d'Angleterre*, publication, traduction, présentation et notes par Christine Ferlampin-Acher, Champion Classiques, Paris, 2007, 289 p.

²⁶ *Ibid.*, p.7.

scholars have used. From the intuitive and rather subjective approaches of the first discussions scholarship has moved towards more selective and elaborate—more scientific—methods of authorship investigation. In this article I propose to present another new approach to the stylistic study of the text—using a pattern recognition method derived from cybernetics. The goal of my research is to prove or disprove the attribution of the romance to Chrétien de Troyes by determining whether syntactic patterns in *Guillaume d'Angleterre* are close to those used by the Champagne poet, based on the conjecture that syntax is the most latent level of language and, consequently, the least exposed to stylistic imitation.

3. Choice of manuscript edition.

The text of *Guillaume d'Angleterre* is preserved in complete version in two manuscripts²⁷. The first, *P*, resides in the Bibliothèque Nationale de France (BnF) under number 375 (old numeration: 6987) and also contains two romances by Chrétien de Troyes: *Cligès* and *Erec et Enide*. Its year of creation is situated about 1288.²⁸ *P* is written in the dialect of Picardy. Apparently this is due to the fact that the copyist was a native of Picardy but copied the romance after its version in the dialect of Île de France.²⁹ The second manuscript, *C*, was discovered in the library of St. John's College Cambridge by Paul Meyer in 1874 and contains *La vie de saint Guillaume d'Angleterre*, *Les Quinze signes de la Fin du Monde*, *La vie de sainte Paule* and other hagiographical stories.³⁰ All the texts were copied at the beginning of the fourteenth century by the same person.³¹ It is a continental version, composed in the Eastern dialect.

As Grégoire Lozinski has shown, the question of identifying the author of the text and the preference of one manuscript to another are interrelated.³² In many cases, the poor effect produced by the text of *Guillaume d'Angleterre* on the reader is not due to the lack of talent of the author, but to the status of the manuscripts, as well as to the changes and mistakes made by copyists. I have based my study on the edition of the romance by Anne Berthelot, published in *Œuvres complètes* in *Bibliothèque de la Pléiade* in 1994.³³ The collection presents a bilingual edition of complete works by Chrétien de Troyes and includes the following texts: *Erec et Enide*, *Cligès*, *Yvain ou le Chevalier au Lion*, *Lancelot ou le Chevalier à la Charrette*, *Le Conte du Graal*, *Philomena*, *Guillaume d'Angleterre* as well as two *chansons courtoises*.

²⁷ Chrétien de Troyes, *Guillaume d'Angleterre : roman du XII^e siècle*, edited by Maurice Wilmette, Paris, Champion, Classiques Français du Moyen Age, N° 55, 1927, p. III-XIV.

²⁸ Paul Meyer, "Compte rendu", in *Romania*, t. 3, 1874, p. 507.

²⁹ Chrétien de Troyes, *Œuvres complètes*, *La Bibliothèque de la Pléiade*, édition publiée sous la direction de Daniel Poirion, Gallimard, Paris, 1994, p. IX.

³⁰ Paul Meyer, op. cit., p. 507.

³¹ *Guillaume d'Angleterre*, traduction par Jean Troitin, Paris, Champion, 1974, p. 12.

³² Grégoire Lozinski, "Compte rendu", *Romania*, t. 54, 1928, p. 275.

³³ I performed a comparison of samples taken from the edition of A. Berthelot that served the basis for my statistical analysis, to corresponding samples from the edition of Ch. Ferlampin Acher, and I found their complete similarity. Thus, my research is compatible with the latest edition of *Guillaume d'Angleterre* based on a different manuscript and it contributes to the assertion that the syntactic level of the language is the less prone to be alternated.

The choice of the *Pléiade* edition as a source of the texts was made for several reasons. First, all the five romances by Chrétien de Troyes (but not *Guillaume d'Angleterre*) are found in the Guiot copy (manuscript *P* with the number BnF 794) which served as a base manuscript for this edition. This manuscript presents the evident advantage of having been copied by the same person, or in the worst case, in the same workshop,³⁴ using identical modes and procedures, something which neutralizes divergences that could appear if the texts were copied by different copyists. The editors chose the manuscript *P8* (number BnF 1450) as a control manuscript: it presented the advantage of having as its content, like the manuscript of Guiot, the whole collection of romances of the Champagne author.

Guillaume d'Angleterre doesn't form part of any of the above-cited manuscripts, and has its own manuscript tradition. The *Pléiade* editor of the romance, Anne Berthelot, chose manuscript *P*, although both manuscripts are of equal value, simply because "C was recently presented to the public thanks to the edition of Holden".³⁵ Nevertheless the general direction of Poirion assured a standard editorial approach applied to all the texts included in the volume. The copy of Guiot procured a homogeneity necessary for processing the text with our chosen statistical method. Finally, the critical edition of each romance provided a clear view "through" the text, and gave us a possibility to consult other variants in case of doubt. The method of pattern recognition is based on the analysis of the syntactic aspect of the text, and we found that the abundance of variants rarely concerned the structure of the text and were often limited to different lexical choices.

By comparing syntactic structures used by "Crestiens" in *Guillaume d'Angleterre* with those that the "father of French courtly romance" used to describe the adventures of his usual protagonists King Arthur, Queen Guinevere and the knights of the Round Table, my aim is to establish how closely distributed the texts are once they are set in a several-dimensional space created by the method of pattern recognition.

4. Using syntax analysis to establish the authorship of a medieval text.

There are several advantages to using a syntax-based methodology for authorship research into a medieval text. The first is closely related to the problem mentioned in the previous part —the choice and status of manuscripts. Every text written before the era of print comes to us as a manuscript, transcribed by a person usually other than its author, normally in a version not reviewed by the author. Scribes would often change some features of the texts, adding or omitting words, and even whole passages. By comparing different versions of the same text we can frequently identify added or omitted parts, and not take them into consideration. More work and attention is

³⁴ As Daniel Poirion points out, "Guiot must have had a permanent workshop, the ink-stand of a merchant, an estal. He was installed in Provins, a town whose market and cultural importance in the county of Champagne, in the environment in which Christian had worked, was considerable"; see Chrétien de Troyes, *op. cit.*, 1994, p. LV.

³⁵ *Ibid.*, p. 1436.

required to compare the lexical structures of the texts, which also could be easily modified by scribes, voluntarily or involuntarily (it is not difficult to imagine a tired clerk substituting by mistake one word by another, or intentionally expressing his own preferences by choosing a more modern word, more suiting his own stylistic feeling or just more popular in that region). On the other hand, syntactic textual structures are more constant and thus less exposed to accidental or deliberate modification by a copyist.

Another advantage is that the syntactic patterns of the text are much less obvious and more difficult to imitate, compared to lexical structures lying on the surface and closely related with the theme, especially since the theme was so explicit in the Middle Ages and required commonly used narrative structures and vocabulary. This means that in order to cause the pattern recognition apparatus to fail, an imaginary imitator, wanting deliberately to project himself as somebody else, would have to disclose those structures in the text of his model in order to reproduce them in his own creations; and the same procedure would have to happen unconsciously in the case of non-deliberate imitation. We can hardly imagine a medieval author proceeding the first way, and the second possibility doesn't seem to be probable, so results issuing from syntactic mathematical tools of attribution are in this sense robust.

4.1. Adjustment of the method for application to the medieval text.

The method of pattern recognition consists in the creation of mathematical models referring to each object (here, a text) on the basis of a certain number of parameters extracted from the text on the basis of the parameters which differ the most, and in the comparison of these models between different texts. The application of the method requires adjustment accordingly to the language of the text and the time of its creation, and Old French is not an exception. A first difficulty concerns the absence of punctuation in the manuscripts. Therefore we exclude all the parameters where the syntactic value of phrase members is determined by punctuation marks.³⁶ One might object that we have used the modern editions of the texts, and relied on their punctuation, but the division by sentences seems to us to be more objective than division according to the sense within the sentence. For the same reason we don't consider any parameters referring to gerundives and participles—for in Old French the distinction between these two classes hasn't yet become clear.

4.2. The description of the objects (texts) using the mathematical apparatus.³⁷

³⁶ For example, number of apposition groups and number of members of apposition groups. Thereafter, to minimize the role of the editor I considered the sentence to be delimited by point or by semicolon; by comparing different editions and versions of the same text I found that often a point and a semicolon are interchangeable.

³⁷ What follows is a brief summary of the methods and results of my approach. For full details see M. Slautina, "L'attribution du roman Guillaume d'Angleterre par la méthode de reconnaissance des formes", thesis defended in 2009 at the University of Caen-Lower Normandy.

The first step in the process of attribution by the method of pattern recognition is a description of the texts-objects in terms of a number of parameters.³⁸ In our case, after the elimination of parameters with a very low frequency of occurrence and parameters not relevant for Old French, the number of parameters is 42. I proceeded to the random extraction of 200 sentences from each object of the first classification, namely *Erec et Enide*, *Cligès* and *Yvain ou Le Chevalier au Lion*. Using the extracted samples I calculated the value of dispersion and the actual dimension of the sample. Afterwards two statistical values were calculated for each object—average arithmetical and standard deviation—that I used in the reduction of the parametric space and in the extraction of those informative parameters which are the most relevant for the differentiation of the texts. This was done using the two-stage Bongard's scheme. The first stage consists in revealing two different classes of parameters—pertinent and non-pertinent—to differentiate objects. The validity of each parameter is determined using the *t criteria of Student* with the threshold value of 1.96 and confidence level equal to $\alpha = 0.05$. If the value of the *t* criteria is higher than the threshold value, the parameter is considered to be informative, otherwise it is excluded from the calculation. After this first stage I got 19 parameters that were pertinent for the object differentiation. The second stage of Bongard's scheme involves reducing the obtained parametric space to the number of informative parameters, with a weak correlation between them and strong correlation with all the other calculated parameters not making part of the final group. This is necessary in order to avoid repeatability of the same information within the group and therefore the actual validity of chosen criteria.

I calculated the efficacy of each of the 19 parameters using this formula:

$$E_i = \frac{\bar{r}_i^{n-m}}{\bar{r}_i^m}, \text{ where } \bar{r}^m \text{ is the average intragroup correlation value, } \bar{r}^{n-m} \text{ is the average intragroup correlation, } n = 42 \text{ and } m = 19. \text{ The criteria of efficacy of 19 parameters is presented in the following table:}^{39}$$

Parameter	\bar{r}^{n-m}	\bar{r}^m	E_i
X02	0.292	0.659	0.442
X03	0.251	0.574	0.436
X04	0.169	0.282	0.599
X06	0.262	0.608	0.431
X07	0.222	0.518	0.428
X08	0.186	0.434	0.429
X11	0.23	0.533	0.432
X16	0.36	0.565	0.637
X17	0.353	0.43	0.821

³⁸ Mikhail A. Marusenko, *Атрибуция анонимных и псевдонимных литературных произведений методами теории распознавания образов [Attribution of Anonymous and Pseudonymous Literary Works by the Method of Pattern Recognition]*, Leningrad, 1990.

³⁹ The numbers of parameters correspond to the initial set of parameters used for different languages including Russian. This initial set contains 54 parameters.

X19	0.214	0.535	0.401
X20	0.322	0.613	0.526
X23	0.269	0.398	0.676
X24	0.351	0.517	0.678
X25	0.222	0.491	0.452
X28	0.224	0.417	0.538
X29	0.351	0.477	0.735
X30	0.213	0.525	0.405
X31	0.164	0.463	0.355
X50	0.217	0.37	0.588

The subset of 19 parameters can be divided into two subsets: a subset of five parameters (X16, X17, X23, X24, X29) whose values efficiency are between 0.637 and 0.821 and a subset of 14 parameters whose values are between efficiency of 0.355 and 0.599. Therefore I manage to extract a subset of five parameters that meet the requirements: having the weakest correlation with the parameters inside the group, having the strongest correlation with the other parameters outside of the group, and therefore being able to distinguish between classes and objects. The model of each romance will be created taking into account all the five parameters simultaneously, which is different from other syntactic methods where each parameter is evaluated separately and then the results are summarized. In other words, the method is holistic rather than analytic.

The five informative parameters are the number of tool words; the number of nouns; the number of prepositions; the number of conjunctions; and the number of words in oblique case. I'll present here an example of extraction of one of the informative parameters from the text, X16, the number of tool words:

En la mer entrent a grant joie,
 Dom Therfes la maistrise avoit,
 Qui del gourenal molt savoit,
Et de le mer et des estoiles (Guillaume, v. 2034-2037).

Total: 10 units

Next, I took the three other texts into consideration. Hereafter all the six objects belonging within our research domain were measured for the five isolated parameters, and their mathematical models were created accordingly. The last part of the attribution process consists in applying deterministic and probabilistic algorithms to attribute the texts to the same or to different authors.

4.3. Deterministic and probabilistic algorithms.

Authorship determination by the pattern recognition method involves two different algorithms of recognition, one deterministic and one probabilistic. The deterministic algorithm is based on the measurement of t criterion of Student: if during the comparison between them the objects present a value less than 1.96 for all 5 measured parameters, the objects are considered to belong to the same class with a probability of 95%. Below I present a table of values of t criterion of Student calculated according to arithmetic average and standard deviation value for each parameter of all the romances of Chrétien de Troyes and *Guillaume d'Angleterre*, where N is the number of treated sentences each time calculated accordingly to the size of the romance:

<i>Cligès</i>			<i>Guillaume</i>			
x_i	σ_i	N	x_i	σ_i	N	t
5.633	4.096	669	4.801	3.273	453	4.824
3.409	2.597	669	2.88	2.121	453	4.499
1.871	1.73	669	1.547	1.414	453	3.870
2.446	2.035	669	2.039	1.818	453	4.004
6.050	4.553	669	5.434	3.891	453	3.070
<i>Yvain</i>			<i>Guillaume</i>			
5.199	3.678	676	4.801	3.273	453	2.335
2.957	2.256	676	2.88	2.121	453	0.662
1.586	1.556	676	1.547	1.414	453	0.477
2.511	1.998	676	2.039	1.818	453	4.662
5.701	4.220	676	5.434	3.891	453	1.336
<i>Erec</i>			<i>Guillaume</i>			
4.477	3.667	737	4.801	3.273	453	1.913
3.120	2.744	737	2.88	2.121	453	2.052
1.588	1.56	737	1.547	1.414	453	0.510
1.682	1.721	737	2.039	1.817	453	3.642
5.325	4.249	737	5.434	3.891	453	0.552
<i>Lancelot</i>			<i>Guillaume</i>			
4.944	3.625	736	4.801	3.273	453	0.845
2.627	2.208	736	2.88	2.121	453	2.225
1.471	1.444	736	1.547	1.414	453	0.952
2.305	1.888	736	2.039	1.817	453	2.678
5.095	4.075	736	5.434	3.891	453	1.721
<i>Perceval</i>			<i>Guillaume</i>			
5.967	4.174	686	4.801	3.273	453	6.765
3.516	2.711	686	2.880	2.121	453	5.390
1.927	1.746	686	1.547	1.414	453	4.549
2.516	1.999	686	2.039	1.817	453	4.713
6.612	5.091	686	5.434	3.891	453	5.825

Applying the deterministic algorithm I fail to attribute the object *Guillaume d'Angleterre*—the romance doesn't present the values of t criterion inferior to the threshold value for all the five parameters together with any of the romances and therefore no close relation with any of the five romances of Chrétien de Troyes was observed. However, these five romances themselves, although their authorship by the poet is not disputed, do not show close links between them: the values of t criterion in their cases vary between 0.291 and 10.009. So I cannot draw any conclusion about attribution from this negative result.

The probabilistic algorithm leads to a more manifest result. It consists in the creation of an initial data matrix and in the transformation of this matrix into a matrix of Euclidian distances between the objects, as presented below:

Romances	$d(X_i, X_j)$					
	Romances					
	<i>Cligès</i>	<i>Yvain</i>	<i>Erec</i>	<i>Lancelot</i>	<i>Perceval</i>	<i>Guillaume</i>
<i>Cligès</i>	0	0.775	1.616	1.476	0.668	1.274
<i>Yvain</i>	0.775	0	1.173	0.772	1.359	0.678
<i>Erec</i>	1.616	1.173	0	0.957	2.2	0.551
<i>Lancelot</i>	1.476	0.772	0.957	0	2.095	0.526
<i>Perceval</i>	0.668	1.359	2.2	2.095	0	1.876
<i>Guillaume</i>	1.274	0.678	0.551	0.526	1.876	0

The next stage is to create a matrix of probability which is calculated on the values of the matrix of Euclidian distances, using a formula that allows one to evaluate the distance between the objects. P_{ji} is the value reflecting the probability of belonging of the object j to the class i, or in our case, the distance between *Guillaume d'Angleterre* and each of the Chrétien de Troyes's romances.

Pairs of romances	P_{ji}
Guillaume / <i>Cligès</i>	0.12
Guillaume / <i>Yvain</i>	0.227
Guillaume / <i>Erec</i>	0.279
Guillaume / <i>Lancelot</i>	0.292
Guillaume / <i>Perceval</i>	0.182

According to this table, in the five-dimensional parametric space *Guillaume d'Angleterre* is situated in the immediate vicinity to Chrétien de Troyes' romances *Lancelot* ($P=0,227$), *Erec* ($P=0,279$) and *Yvain* ($P=0,227$). The two other romances participating in the research, *Le Conte du Graal* and *Cligès* were situated further from the borders of the

a priori class formed by the above-mentioned texts (respectively $P = 0,182$ and $0,12$). For a graphical representation of the results see Graph 1.

What this analysis demonstrates, in simple terms, is that the models of four romances, built on their syntactic patterns including number of tool words, number of nouns, number of prepositions, number of conjunctions and number of words in oblique case, belong to the same class of objects, and when I integrate the object *Guillaume* in the same class as the objects *Lancelot*, *Erec* and *Yvain*, I state its qualitative homogeneity and amelioration of the condensation of the class—the necessary condition of a cluster. On the other hand, if I include *Cligès* and *Perceval*, the homogeneity of the class deteriorates.

This might be explained by the fact that the copy of Guiot does not have the same value for all the texts; for example, in the case of *Le Conte du Graal*, it is considered furthest from the original, which is confirmed by comparisons with other lessons from other manuscripts. Another possible explanation is that this text is commonly considered to be the last work of the writer, most critics agreeing on the fact that the novel remained unfinished because of its author's death. The explanation of the remoteness of *Cligès* is more problematic—is it possible to conjecture that this work equally belongs to the late period of Chrétien's literary activity, even if the romance is more commonly admitted to be second text of his production, after *Erec et Enide*? In his recent article dedicated to the prologue of *Cligès* Levilson Reis, basing his opinion on the works of Paul Zumthor, Noris Lacy and Michelle Freeman, argues that the enumeration of texts written by Chrétien by the time of composition of the prologue of *Cligès* has a conceptual rather than chronological meaning—helping to “situate *Cligès*”—and thus the tradition of placing *Cligès* immediately after *Erec* and before *Lancelot* and *Yvain* might be reconsidered (see Graph 1).⁴⁰

5. Conclusion.

In this article I have presented for the first time an attempt to apply the method of pattern recognition to determine the authorship of a text written in Old French. The results of my work demonstrate that according to strictly established and statistically relevant rules, this method is applicable not only to compositions by modern authors, but also to older texts that have survived only in copied versions. Similarly, I consider that my extension of the method of pattern recognition to the field of Old French has given better results than the older methods I reviewed at the beginning of this article, and can justify further research in the same general domain, medieval literature being rich in examples of texts with doubtful attribution, the resolution of which is essential in literary and historical studies.

⁴⁰ Levilson C. Reis, "The Paratext to Chrétien de Troyes's *Cligès*: a Reappraisal of the Question of Authorship and Readership in the Prologue", in *French Studies*, 2010, vol. 65 (1), p. 8.

Graph 1:

