

Words from BOnonia Legal Corpus

R. ROSSINI FAVRETTI, F. TAMBURINI AND E. MARTELLI¹
CILTA, University of Bologna

*The analysis of special multilingual corpora is still in its infancy but it may serve a particularly important role for the directions it offers both in cross-linguistic investigation and in the selection of the most typical features of text types and genres. To exemplify the information which can be obtained from corpus evidence, the paper reports on an on-going corpus-driven research project, named Bononia Legal Corpus or BOLC. The main aim of BOLC is to build multilingual machine readable law corpora. Data are at present limited to English and Italian, but an extension is envisaged to include other languages. Before the first sample, a preliminary pilot corpus was constructed to consider European legislation and create a conceptual framework to be used as a first-level experience. In the paper, sections 2 and 3 describe the corpus design and formatting as well as the corpus access tools. Sections 4 and 5 discuss two case studies and analyse two semantic areas which can be seen as two ends of the same variational continuum. At one end, we consider the words *contratto* and *contract*, which through the extension of international transactions and circulation, may be supposed to have acquired transnational traits. At the other, we focus on a semantic area which may be expected to present translation problems for the differences existing in the two socio-institutional systems. Reference is made to the English words *tax* and *duty* and to the Italian words *tassa* and *imposta*.*

KEYWORDS: Corpus linguistics, corpus data processing, lexicography, semantics, cross-linguistic comparison.

1. Introduction

The use of computer-based text corpora can be considered one of the most significant developments in linguistic research in the last decade. Text processing has opened wide perspectives in the investigation of data for scientific purposes. It has become a major concern to approach linguistic data through large corpora of naturally occurring language, attaining insights into different levels of language description. On the one hand, the approach has been facilitated by the developments in hardware technology and by on-line access to textual resources. On the other, it has taken advantage of computational techniques for the retrieval and statistical processing of the data.

¹R. Rossini Favretti took charge of sections 1, 4, 5 and 6., F. Tamburini and E. Martelli took charge of sections 2 and 3.

Corpus linguistics has had an important impact on different aspects of linguistic research and statistical tabulation has proved to be a basic starting point not only for quantitative but also for qualitative analysis of different types of language. A high number of general corpora were constructed and relevant results have been obtained. In our opinion, anyhow, corpus evidence may serve a particularly important role in the analysis of special corpora for the directions it offers in the investigation of large samples of texts and in the selection of the most typical features of text types and genres.

The paper reports on an ongoing corpus-driven research project carried out at the University of Bologna. The main aim of the project - named Bononia Legal Corpus, or BOLC - is to build multilingual comparable machine-readable law corpora. It is an interdisciplinary project and John Sinclair has played a crucial role as consultant. Work was begun in 1997 and, if everything goes according to plan, carrying out the project will take three years - 1997-1999. Data are at present limited to English and Italian, but an extension is envisaged to include other languages. As to the size of the corpus, we set 10 million words as the smallest target for each component.

English and Italian legal texts were chosen as representative of two different legal systems and of differences existing between the common law system developed in England and the civil law system, based on the Roman law, developed in Italy. Before the first sample, a preliminary pilot corpus was constructed to consider European legislation for the transnational dimension which is implied in the coexistence and cooperation of different nationalities. It was directed at creating a conceptual framework to be used as a first-level reference. We chose to refer to secondary Community legislation and in particular to "Directives" and "Judgments" as they may be implemented by domestic legislation and may produce direct legal effects in member states. They are seen as text types on either side of the border between parallel² and comparable³ corpora. As the texts are to be representative of contemporary legal language the documents chosen were issued in the period 1968-1995.

Reviewing briefly, the research is aimed at providing contrastive information on meaning and usage to guide lexicon builders and at indicating the standards of accuracy and detail required of future lexicons to be effective tools for translation and other applications.

In this paper, sections 2 and 3 describe the corpus design and formatting as well as the tools used to access corpus data. Sections 4 and 5 discuss two case studies on the basis of the analysis carried out in the pilot corpus now available - about 18 m.w. . We consider two semantic areas which can be seen as two ends of the same variational continuum. At one end, we will consider the English word "contract" and the Italian word "contratto" which, through the extension of international transactions and circulation, may be supposed to have acquired transnational traits. On the other, we will

² A "parallel corpus" has been described as "a bilingual or multilingual corpus that contains one set of texts in two or more languages"(Teubert, 1996, 245). According to Teubert, it may contain 1) only texts originally written in language A and their translations into languages B (and C...); 2) an equal amount of texts originally written in languages A and B and their respective translations; or 3) only translations of texts into the languages A, B and C, whereas the texts were originally written in language Z .

³ The term "comparable" is used to describe corpora in two or more languages that have a similar composition and can be compared because of their common features.

focus on a semantic area which may be supposed to present translation problems for the differences existing between the two socio-institutional systems. Reference will be made to the English words "tax" and "duty" and to the Italian words "tassa" and "imposta" .

2. Corpus design and formatting

The BoLC pilot corpus consists entirely of European Community documents, mainly directives and judgments. The documents exist in English and Italian and cover the production from the founding of the European Community to March 1995 for the Italian documents and to July 1996 for the English documents. It is important to underline that the Italian documents are a translation of the English ones, because the European Community draws up its original documentation only in English and French.

We collected approximately one hundred and ten megabytes of electronic text for each language, divided as shown below:

2,232 Directives: 6,500,000 words,
1,798 *Direttive*: 5,800,000 words,
4,472 Judgments: 13,700,000 words,
4,471 *Sentenze*: 12,300,000 words.

The retrieved documentation was not directly usable because there was a lot of additional information mixed with the essential text and a lot of orthographic errors. So a great deal of work was required to eliminate from the documents all that was unnecessary and inessential, and to correct the mistakes. A lot of reference tags, multiple blanks between words, blanks between words and punctuation marks were removed to standardise the document formatting and to save space. The documents were coded in SGML ISO-Latin-1 to make the corpus platform independent. The problem was that in the original documents there were a lot of characters, especially accents in Italian, which are correctly displayed in a DOS computer, but not on different ones. The SGML coding is an international standard for multilingual documents, correctly handled by different computers.

In the earlier Italian documents there were wrongly written words, some others without accents and so on. We solved this problem by comparing each word with an electronic dictionary, augmented with all the Italian verb conjugations, inserting all the requested accents and fixing most of the remaining errors.

Finally the single documents were joined together in four subcorpora and then indexed to be correctly handled by the corpus access tools.

3 Corpus access tools

3.1 *Corpus data retrieval*

Nowadays there is an increasing need for large corpora, both to investigate changes in everyday language - such as “monitor corpora”, that foresee no finite size but a flow of information and linguistic evidence filtered through devices, to create an exact picture of the real up-to-date language (Sinclair 1991) - and to analyse extremely specialised linguistic features. In order to manage this amount of data, we need adequate computational procedures that have to be *general* - they have to accept different

approaches to mark-up, tokenisation, languages, etc. - *flexible* - they must allow corpus maintenance and adaptation - *user friendly*, and, last but not least, they have to be *extremely fast*. In response to these needs O.Mason (1996) has devised CUE (Corpus Universal Examiner), a set of computer programs able to address all the requirements of a modern corpus retrieval application. The first version of CUE was written in C++ for UNIX systems, using the publicly available library Xforms (Zhao and Overmars 1995, Reichard and Johnson 1996) for the interface design. It involves complex indexing schemes (inverted index), fast procedures for the retrieval and access of data and compression methods (Huffman coding) to reduce the amount of space needed to store the corpora. The main problem with this application was that it followed the standalone application paradigm. This meant that only the workstation that stored the corpora would have immediate access to them. Even if a complete Networked File System were provided the application would run only on UNIX machines.

When we started the BOLC project it was immediately clear that having only one station with corpus access did not meet our needs and we had to provide a different access method for users. The decision was to transform the standalone version of CUE into a client-server application, in such a way that the server machine can provide corpus access across our Local Area Network. Moreover, we had to address a different problem, the multi-standard nature of our client workstation. At CILTA we currently have Windows based PCs, Macintoshes and UNIX workstations. It was not conceivable to develop and maintain a different client application for each kind of operating system/hardware platform pair. The natural, and unique, solution to such problem was to develop the CUE client side in Java, obtaining, in theory, complete portability among different systems without any further effort.

Figure 1 shows the scheme of the new version of CUE (called JCUE), developed at CILTA.

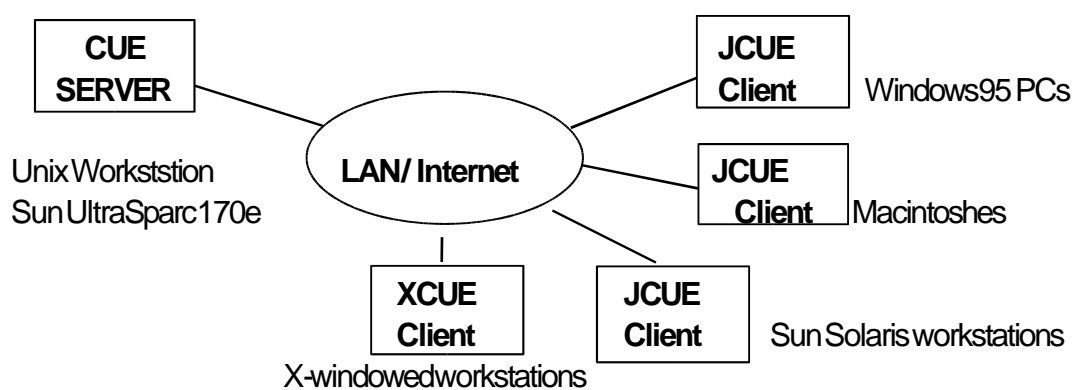


Fig 1. Client/Server structure of JCUE, developed at CILTA.

The server side was derived from the original CUE release. It is written in C++ and runs on a Sun UltraSparc 170e with 96MB of memory and 5GB of disk space supporting the Solaris 2.5.1 operating system. It was implemented following the concurrent server model, so that it can accept multiple queries from different client machines at the same time. Once a new client makes a request to activate the service, a

new copy of the server program is created; it remains active once the client closes the connection. It is important to note that, for security reasons, the client has to provide authentication - as a legal JCUE client program - and the user, who is trying to access this service, has to provide passwords. In this way we can restrict the use of some corpora to particular users or research teams.

The most complex work was to divide the standalone application into a server side and a client side, providing a complete set of operations needed to retrieve data from the network. We developed a scheme similar to Remote Procedure Call technique, building a client-and-server-module interface to the network communication protocol. Fig 2 outlines the methods.

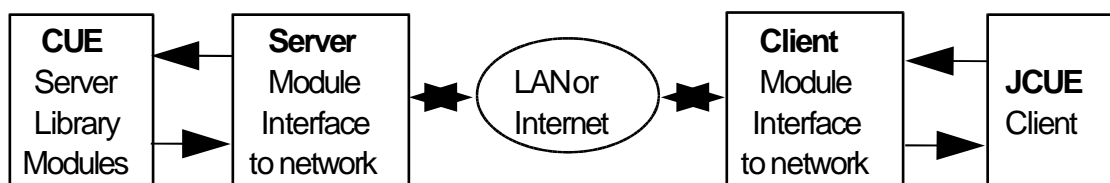


Fig 2. Communication structure for JCUE package.

These modules transform the request and the data from the client side in string codes that are sent across the network using the standard BSD socket support. Using a similar scheme, they transform the data retrieved by the server in a similar way and send it back to the client.

The client side was completely redesigned using Java (version 1.0.2), and is currently working on Windows 95/NT PCs, Macintoshes, Sun-Solaris UNIX workstations. We faced a number of problems using Java, mainly due to the differences among the implementation of the Java runtime machine on different architectures. This is why we decided to develop the client in the first, widely implemented, version of Java. We also developed an X-Window version of the client for UNIX machines, directly derived from the original CUE package.

3.2 Source document extraction

For an in-depth analysis of parallel corpora it is often not sufficient to examine only the concordances produced using a retrieval procedure. Sometimes, in order to clarify the relationship among words from different languages, it is necessary to examine the entire document that contains a determinate concordance, even if features that furnish the extended concordance context are available. Moreover, this kind of analysis is often carried out using separate programs that *align* parallel document texts.

In order to satisfy these needs, we developed a system for document identification and a separate client-server application for the document retrieval. This application, that we called Corpus Document Extractor (JCDE), behaves in a similar way as JCUE package. A server, written in C++, runs on the station that contains the corpus data, while a Java client, that communicates with the server across the network, interfaces the document retrieval procedure from every remote station (Windows 95/NT PCs, Macintoshes, UNIX workstations). Using this client/server application the user can

retrieve the documents contained in the corpora, specifying only the document identification string.

4. The terms "contratto" and "contract": translation equivalences

To illustrate the information which can be obtained about the syntactic and semantic structures of the terms under investigation, as an example, the term "contratto" was selected from the Italian subcorpus and used as the search node.

The selection of the term was determined by the relevance of the contract as a legal device. The contract, it has been argued, may be considered as the legal cornerstone of all transactions in business and consumer life. The law of contract is deeply embedded in the business practices of different countries. Different legal systems may vary substantially on a number of matters owing to historical, institutional or commercial reasons, but in recent times, with the rapid expansion of trade and business, attempts have been made to limit the effect of dissimilarities in the contract law of different legal systems. A process of "internationalization" may be assumed, in spite of the deep-rooted divergencies still existing between the systems of common law and civil law.

To identify the collocates of the term "contratto" the concordances were automatically selected from 4,642 citations:

di un anticipo sull' aiuto relativo al contratto , anticipo che le veniva versato dalla
forniti o non siano comunque conformi al contratto di fornitura . 2 . quando : a) per l'
cisione finale sull' aggiudicazione del contratto , sono prese da detto stato . le contr
bouyer) , relativa alla risoluzione del contratto ed alla condanna al risarcimento dei d
atto loro perdere l' aggiudicazione del contratto d' appalto per la costruzione dell'
triennio successivo alla conclusione del contratto d' appalto iniziale ; h) quando , ec
to del danno o chieda la risoluzione del contratto per inadempimento della controparte ,
auzione a garanzia dell' esecuzione del contratto garantito) condizioni particolari del
in detta tabella . la caratteristica del contratto di agente ausiliario e la precarietà q
impresa a seguito della risoluzione del contratto di locazione - vendita mediante pronun
simo di due anni dopo l' estinzione del contratto . 4 . il presente articolo lascia impr
 , in sostanza , che la comunicazione del contratto Statoil non e' " necessaria " , poich
lla commissione nell' inadempimento del contratto per una colpa commessa all' atto de
pagamento di diverse somme in forza del contratto di lavoro o a causa della sua disdetta
rattotto giorni dopo la stipulazione del contratto in questione " . 20 gli artt . 17 - 25
ncanze constatate nell' adempimento del contratto non siano imputabili ne' a colpa loro
le si applichino fino alla scadenza del contratto . Se necessario e' possibile assumere
nvenuto soltanto dopo la conclusione del contratto di ammasso . 2) l' operazione che se
itti e gli obblighi che ha in virtu' del contratto d' agenzia . articolo 19 le parti non
erprete o esecutore contemplato da detto contratto) applicato sul risone prodotto in ita
ettore l' onere pecuniario (diritto di contratto di lavoro e' quella che caratterizza
prendere in considerazione in materia di contratto di lavoro , compreso il mantenimento d
mantenimento dei diritti connessi con il contratto di lavoro , le disposizioni dello stat
tutato e relative all' esecuzione di un contratto scritto di concessione esclusiva di ve
ttributiva di competenza contenuta in un contratto di ammasso di formaggi e disciplinata
gomento secondo cui la conclusione di un contratto di appalto di lavori pubblici finanzia
o estromesse dall' aggiudicazione di un contratto di compravendita di latte intero norma
zione , per il 30 settembre 1978 , di un contratto di fornitura di mangimi stipulato tra
civile concernente l' esecuzione d' un contratto di concessione di licenza , secondo la
e sub 1 : se la clausola contenuta in un contratto di concessione di licenza , secondo la

As a following step the term "contract" was selected from the English subcorpus and these concordances were automatically selected from 5,449 citations:

posts . An important characteristic of a contract for the employment of auxiliary staff
centres with which they have concluded a contract for the supply of animals or semen . 5
invited to state first of all whether " a contract for the supply of beer concluded befor
part of that training takes place under a contract of apprenticeship concluded under the

posts . An important characteristic of a contract for the employment of auxiliary staff respect of obligations which arose from a contract of employment or an employment relationship with the flexon - italia undertaking a contract for the cleaning of the establishment Conclusion and termination of the agency contract Article 13 1 . Each party shall be ent a transferor resulting from an employment contract or employment relationship and arising following entry into force of the export contract , shall be the condition precedent to of contract : 4 . Criteria for award of contract : 5 . Number of tenders received : 6 . e concerning indemnity for termination of contract between the principal and the commerci nt precluded on the grounds of freedom of contract of the parties to the Collective Agree ed the public works at issue by a private contract and had failed to publish a notice of ing authorities who have awarded a public contract or have held a design contest shall se and , if necessary , adjust the research contract to the new situation with the applican rformance by the other party to the sales contract under which the goods were to be expor mine a counterclaim arising from the same contract or facts on which the original claim w tract . 7 . Criteria for the award of the contract . 8 . Other information . 9 . Date of ate , the agency or branch concluding the contract is situated (a) 3 . The address of t in such a list in the state awarding the contract may be required of contractors establi k ; (d) the date of commencement of the contract or employment relationship ; (e) in before the date of the conclusion of the contract . 3 for the 1971 / 72 wine - growing y be considered suitable to tender for the contract in question . However , such a mention tion for admittance to participate in the contract that , during the three previous years e of his rights and obligations under the contract without the franchisor ' s approval with Belgian law , the dissolution of the contract by the court , on the ground of the gr be required to do so if it is awarded the contract , to the extent that this change is ne uent proof of Fiat ' s strong position in contract negotiations . (721 et seq .) . 146

If we begin by examining the environment of the term "contract", we notice that "contract" appears 1) as a headword , 2) as a modifier of a noun group or 3) as a single-word term, often preceded by a determiner.

Let us consider the first position to the left of the node (designated N-1). We find two kinds of collocates: grammar words and full lexical words. Both in Italian and in English concordances we notice a high occurrence of the article - both definite and indefinite - often preceded by a preposition, in N-2 position. "Of" and "di" dominate the pattern. In each of the tables if we look at N-3 position we notice the occurrence of a noun. A regular pattern can be identified in the following noun groups where processes inherent in the commencement, performance and conclusion of the contract are expressed:

award	of (the) contract	aggiudicazione	del contratto
breach		inadempimento	
conclusion		conclusione	
commencement		inizio	
dissolution		scioglimento	
execution		esecuzione	
performance		adempimento	
publication		pubblicazione	
rescission		estinzione	
signature		firma	
stipulation		stipula, stipulazione	
suspension		sospensione	
termination		risoluzione	

A noun group emerges as particularly relevant:

noun + *di* [+ determiner] + *contratto*

noun + *of* [+ determiner] + *contract*

where the noun is a derived nominal and the subjective value of terms denoting the contract is constant:

1. (a) la conclusione del contratto
1. (b) il contratto è concluso

2. (a) the conclusion of the contract
2. (b) the contract is concluded

In the collocations provided in the tables a number of equivalences may be identified in the lexicalization of the contract procedures, but a difference emerges, even from a superficial glance, in the conceptual extension of the terms "contratto" and "contract". In a number of concordances, corpus evidence suggests two different senses for "contract" which have their translation equivalents, in Italian, in 1) "contratto" and 2) "contratto d'appalto". A striking feature in the tables is that various kinds of lexically specific information is associated with "contract" in:

2.(a) the conclusion of the contract

and in :

3. the award of the contract

The nature of the contract, in its most salient and typical components, is strictly tied to the collocate, particularly, in 3, to the word "award". "Award" is a far more important collocate (610) in English than "aggiudicare" (55) and "aggiudicazione" (7) are in Italian. To illustrate this point let us consider the following citations selected automatically from our corpus:

the conclusion of a contract following its award , the powers of the body responsible fo
of the grounds on which it decided not to award a contract in respect of which a prior
te . 2 . Where the contracting authorities award a contract by restricted procedure , th
of the grounds on which it decided not to award a contract in respect of which a prior
s relating to the contract provide for its award at the lowest price tendered , the cont
: - either require the concessionaire to award contracts representing a minimum of 30
2 . Number of contracts awarded (where an award has been split between more than one su
ized as part of a procedure leading to the award of a service contract the estimated val
uests to participate in procedures for the award of contracts may be made by letter , by
ption . CPC reference number . 4 . Date of award of the contract . 5 . Criteria for awar
r (Article 16 m) : 13 . Criteria for the award of the contract . Criteria other than t
ember 1976 coordinating procedures for the award of public supply contracts (6) , as l
om the scope of the law procedures for the award of public works contracts other than by
h Article 40 , information relating to the award of contracts . 3 . As regards individua
cerning coordination of procedures for the award of public works contracts (89 / 440 /
icles 25 And 26 (d) the criteria for the award of the contract if these are not given
oordination of national procedures for the award of public supply contracts ; Whereas su
lection of suppliers or contractors and of award of contracts , contracting entities may
VISIONS Article 28 For the purposes of the award of public contracts by the contracting
tors have a fair opportunity to secure the award of contracts , but does not contain any
tation . Article 7 For the purposes of the award of public contracts by the contracting
the commencement of the procedures of the award of the contract (s) (if known) . 4
ement has been committed during a contract award procedure falling within the scope of D
the contracting authority : 2 . (a) The award procedure chosen : (b) Form of the co

ner participating in the relevant contract award procedure the opportunity to make repre the contracting authority : 2 . (a) The award procedure chosen : (b) Where applicab s of the contracting authority . 2 . (a) Award procedure chosen . (b) Where applicab nting that law as regards : (a) contract award procedures falling within the scope of he tenders before deciding to whom it will award the contract . For this purpose it shall than a contracting authority , who wish to award works contracts to a third party within

"Contract" may occupy different positions in the verbal co-text of "award", but it is always present in its role structure.

At this point, it is worthwhile considering the patterns in both languages. Let us examine the concordance of the limited examples of "aggiudicazione" in Italian:

delle nuove forme contrattuali di aggiudicazione degli appalti e introdurre crit opo di coordinare le procedure di aggiudicazione dei contratti di appalto di lav lavori da dare in appalto e l ' aggiudicazione del contratto sono due operazio l . Laddove il criterio per l ' aggiudicazione del contratto sia quello dell ' di un contratto in seguito all ' aggiudicazione dell ' appalto , i poteri dell di un contratto in seguito all ' aggiudicazione dell ' appalto , i poteri dell di appalti ; considerando che l ' aggiudicazione di contratti relativi a determi

In Italian "aggiudicazione" and "appalto" are important collocates of the term "contratto" but in a number of examples they occur without "contratto" as a collocate. As far as we can ascertain in our corpus, "contratto" and "appalto" are not necessarily "mutually expectant words". The following concordance of "appalto", automatically selected from 728 citations, may illustrate this point:

er le forniture cui si riferisce l ' appalto , relativo agli ultimi tre esercizi successivo alla conclusione dell ' appalto iniziale . 4 . In tutti gli altri c UARE TALE TRASFORMAZIONE QUALORA L ' APPALTO GLI VENGA AGGIUDICATO . ARTICOLO 22 calcolo del valore di stima dell ' appalto : - nell ' ipotesi di appalti una d alitativa e di aggiudicazione dell ' appalto e che esse non prevedono la possibi alcolo dell ' importo stimato dell ' appalto è : - se trattasi di appalto di dur al quale sarà stato aggiudicato l ' appalto : 6 . a) Data limite di ricezione he cos tituiranno l ' oggetto dell ' appalto ; b) l ' avviso deve indicare che seguito all ' aggiudicazione dell ' appalto , i poteri dell ' organo responsabi per partecipare ad una procedura d ' appalto o ad un concorso di progettazione ; VERSIA SORTA DA UN BANDO DI GARA D ' APPALTO DELL ' ADMINISTRATION DES PONTS ET purché le condizioni iniziali dell ' appalto non siano sostanzialmente modificat separabili dall ' esecuzione dell ' appalto iniziale , siano strettamente neces ' AGGIUDICAZIONE DEL CONTRATTO D ' APPALTO PER LA COSTRUZIONE DELL ' ISTITUTO . c) Eventualmente , forma dell ' appalto che è oggetto della gara . 3 . a) NECESSARIE NEL CORSO DELLA GARA D ' APPALTO , COMPRESA LA DECISIONE FINALE SULL fferenti e l ' aggiudicazione dell ' appalto possano aver luogo simultaneamente MPRESE CHE PARTECIPANO ALLE GARE D ' APPALTO O ALLE QUALI SONO AGGIUDICATI APPAL lo di gara relativo al contratto d ' appalto n . 4 del progetto relativo all ' a IONE , A TRATTATIVA PRIVATA , DELL ' APPALTO PER LA REALIZZAZIONE DELL ' IMPIANT ATA IN GRADO DI AGGIUDICARE UN NUOVO APPALTO . PER I MOTIVI GIÀ ESPOSTI IN PRECE CCIANO O MENO PARTE INTEGRANTE DI UN APPALTO DI LAVORI PUBBLICI . 3 . L ' ARTICO usole contrattuali di un determinato appalto , di prescrizioni tecniche che menz ori all ' impresa titolare del primo appalto , a condizione che i nuovi lavori s ditore che desideri partecipare a un appalto pubblico di lavori può essere invit ente - Riserva di una frazione di un appalto pubblico alle imprese situate in un catrici e che intendono stipulare un appalto di lavori con un terzo , ai sensi d le amministrazioni aggiudichino un appalto mediante procedura negoziata second onsiderare un accordo quadro come un appalto ai sensi dell ' articolo 1 , paragr di automazione del gioco del lotto ° Appalto non riguardante attività che implic

All these patterns:

4. l'aggiudicazione del contratto d'appalto
5. l'aggiudicazione degli appalti / dell'appalto
6. l'aggiudicazione del contratto

find their translation equivalence in:

3. the award of the contract

In English it is the process expressed by the verb "award" which is associated with the peculiar typology of contract 2. What can be argued, in the present connexion, is the fact that in all the English examples of the corpus it is in the collocates such as "award" and tender that we find the lexical information which is associated, in Italian, with "contratto d'appalto" or "appalto".

A second notable feature which emerges in the comparative analysis of the tables of "contratto" and "contract" is the way in which the contract type is specified through pre-modification (N-1) in English and post-modification (N+1 and N+2) in Italian :

7. agency contract

8. contratto d'agenzia

Examples of post-modification may be found also in the English subcorpus, but pre-nominal modification prevails in English whereas post-nominal modification prevails in Italian.

If we look at the syntactic environments of the words "contratto" and "contract", a further difference between the syntactic structures of the two languages is illustrated by the class shift taking place when "contract" occurs as modifier:

9. contract negotiations

10. negoziazioni contrattuali

The word "contrattuale" has a high occurrence (490) in Italian examples and "contract" is its translation equivalent in English:

aria del dipendente di ruolo e quella , contrattuale , dell ' agente temporaneo , una di
vi sia cambiamento , dovuto a cessione contrattuale o a fusione , della persona fisica
rsi da quelli operati mediante cessione contrattuale oppure mediante fusione , quest ' u
ce a carico della commissione una colpa contrattuale di cui essa deve rispondere . tale
errori o carenze nel suo comportamento contrattuale , come un ritardo nell ' approvazio
embri in fatto di responsabilita' extra contrattuale . quanto al problema della prova de
ronunziarsi sulla responsabilita' extra contrattuale della comunita . 4 . la constatazio
. 5 in materia di responsabilita' extra contrattuale , il trattato assoggetta la comunit
e non attribuisca importanza alla forma contrattuale - acquisto o leasing - nemmeno nel
re 1968 - competenze speciali - materia contrattuale - concessione esclusiva - lite fra
mpimento dell ' obbligazione in materia contrattuale . 19 e ' vero che questa norma non
guita ... ' . 9 la nozione di materia contrattuale serve quindi di criterio per delimi
a parte della prima dell ' obbligazione contrattuale di consegnare alla Rewe - zentral 2
voro , al di fuori di qualsiasi obbligo contrattuale , conceda speciali agevolazioni di
n ' assicurazione avente base puramente contrattuale non rientra quindi , razione materi
duttori , nonche' in materia di diritto contrattuale . qualsiasi disposizione contrattua
uto e che non si ricollega alla materia contrattuale di cui all ' art . 5 , punto 1) .
ha mai assunto alcun obbligo di natura contrattuale nei confronti del subacquirente ste
azioni) , che lo statuto ha una natura contrattuale e che , percio' , una clausola attr
iudicata la liberta' della negoziazione contrattuale dei diritti sancita dalla presente
a questione sub 1 : se l ' obbligazione contrattuale , secondo la quale il concessionari
ere i) , da un lato , nella sua prassi contrattuale , imposto alle sue controparti un d
* di nave * in tonnellate * del prezzo contrattuale (1) (1) l ' equivalente - sove
o un peso pari al 90 % del quantitativo contrattuale , a prescindere dal fatto che i pez
ale , senza raggiungere il quantitativo contrattuale , l ' importo dell ' aiuto viene ri

This may be traced back to the different formation of noun groups in the two languages. In English most noun groups consist of two or more nouns. In Italian, they predominantly consist of a noun either preceded or followed by one or more adjectives. This can have an important bearing on our analysis of right and left collocates.

If we go on in our analysis and consider the first position to the right of the node (N+1), we find prepositions as predominant collocates. The preposition *of* (821) and the preposition *di* (1,386) prevail, followed by a noun in N+2 position:

contratto + *di* + noun
contract + *of* + noun

Another notable feature, in English, is the occurrence of the preposition *for* (217) when the noun is preceded by the definite article. When *for* is associated with a determiner and a noun, the noun is usually qualified by a prepositional phrase:

contract + *for* + determiner + noun + *of* + noun

A constant distinction is drawn between phrases like:

11. a contract of employment

and phrases like:

12. a contract for the employment of auxiliary staff

Such distinction has no equivalent in Italian:

un contratto + *di* + noun [+ *di* + noun]

In the cross-language analysis, we can say that syntactic differences play a more important role than lexico-semantic ones. It remains to be seen whether these results have a general value or are limited to the terms under scrutiny.

5. Translation equivalents of the terms "tax" and "duty"

5.1. The term "tax": what the English subcorpus shows

To exemplify a situation where cross-language equivalence cannot be assumed we will refer to the tax law and analyse, as a second case study, the word "tax". Through the word "tax" a situation is referred to which can be considered common both to England and to Italy and can be assumed to apply, with the extension of our corpus, to other European countries as well. In all countries, taxes are levied on income and expenditure by central and local governments, but different categories are employed in their

definitions. It is our hypothesis that some of the main categories may emerge from interlinguistic comparison.

As a first step, we will consider the following concordance of the word "tax", selected automatically from our corpus where there are 7,722 citations altogether:

se gave rise , for the purposes of turnover tax , to a new immovable property comprising the purposes of the rules on value - added tax , by agreement with one of his employees ied out as long ago as 1967 , only turnover tax , which at that time was applicable at t necessary steps to permit the remission of tax , in accordance with the procedures refe fic rates of the Portuguese motor - vehicle tax , which increases sharply as from a spec over taxes - common system of value - added tax - duties or charges which cannot be char o apply section 10 (2) of the value added tax act 1972 , which reduces the taxable amo ners for the special purposes of the income tax acts , hereby rules : Community law proh ulfilled a Member State may not refuse that tax advantage on the basis of supplementary national legislation for qualifying for the tax advantage in question . ISSUE 1 in the t that , by granting exemptions from turnover tax and excise duties in respect of the impo plementing the common system of value added tax and amending Directive 77 / 388 / EEC (principle , goods acquired free of turnover tax and excise duties in the course of Intra IVE : Article 1 1 . Exemption from turnover tax and excise duty on imports shall apply , criteria laid down by law , which give the tax authorities no discretion and make no di mely that a system of road tax in which one tax band comprises more power - ratings for proceedings instituted by H . Lennartz , a tax consultant in Munich , concerning the re ver , the Commission has not challenged the tax differential between sparkling wines tax iance with the rule that there should be no tax discrimination - (EEC Treaty , art . 95 onditions , be justified in an area such as tax law , it must be observed in this Case t hat winding - up entails in company law and tax law . The legislation of other States pe tation of the programme of harmonization of tax legislation pursuant to Article 99 of th f , he shall be entitled to deduct from his tax liability the value added tax due or pai ose of his business , where the value added tax on the goods in question or the componen e of taxation Whereas a Community system of tax reductions on imports has proved necessa laid down by Member States until Community tax rules are adopted . The exemption may be , the chargeable event shall occur and the tax shall become chargeable at the time when ccordance with the cumulative multi - stage tax system has constantly given rise to diff authorities of the Member States where the tax warehouse is authorized ; (b) comply w er States of a common system of value added tax Whereas a system of value added tax achi

On inspecting the concordances we observe that "tax" tends to occur either followed or preceded by a noun, or a noun group. Like "contract", it occurs 1) as a modifier, 2) as a headword, and 3) as a single word term.

In a particularly high number of examples it occurs as a modifier in a noun group. As its top ten collocates, in N+1 position, we find:

provisions (337)
system (196)
purposes (165)
authorities (132)
burden (101)
legislation (97)
advantages (93)
arrangements (81)
exemptions (65)
exemption (58)

In the examples where the term "tax" occurs as a headword, it is associated with pre-nominal (N-1) or post-nominal (N+1) modification. N-1 position may be occupied:

- by a noun

turnover (605)
income (102)

- by an -ed modifier

value-added (664)

- by an -ing form

withholding (12)

In the examples where the word "tax" is not associated with pre-modification, N-1 position is occupied:

- by a preposition

of (588)
for (165)
to (157)
from (71)

- by an article

the (1,294)
a (324)

On the right, where a noun does not occur in N+1 position, the position is often occupied by a preposition and "tax" is qualified by a prepositional phrase:

on consumption (49)

The occurrence of "tax" without modification tends to concentrate in instances where the term is either preceded or followed by a comma or by connectives:

duty and tax
turnover tax and excise duty

The examples suggest that "tax", in its singular form, presents three different senses:

- 1) a general, indefinite one, in the first instances, when followed by a noun and used as modifier;
- 2) a general collective one, in the second group of instances, when it is not associated either with post-modification or with pre-modification;
- 3) a specific one, when it is preceded by a modifier in N-1 position.

There is a hyponymic relation between 3 and 2, which may be exemplified by such pairs as "turnover tax" and "tax".

5.2. The term "duty"

In the concordance of "tax", "duty" appears as a significant collocate. "Duty" collocates with "tax", but the lexical environments of the two words is different. Their most prominent collocates do not overlap, as the concordance below, automatically selected from 5,705 citations, illustrates:

basis adopted for the imposition of excise duty . 4 the appeal lodged by gb - inno -
arge having equivalent effect to a customs duty , and The application of any quantitativ
oods other than products subject to excise duty , paragraph 1 shall not apply to supplie
arge having equivalent effect to a customs duty , contrary to Articles 12 et seq . of th
nt the Commission objections regarding the duty - free importation of the instrument or
ic drinks , the real value of the rates of duty and the wider objectives of the Treaty .
oleum products , both net and inclusive of duty and tax _ the estimated average gross ex
selling prices , both net and inclusive of duty and tax , whether published or not , for
ortional excise duty , the specific excise duty and the turnover tax levied on these cig
uty and the sum of the proportional excise duty and the turnover tax , in such a way tha
e having an effect equivalent to a customs duty but is in reality intended to offset exa
duty which may be : - either an ad valorem duty calculated on the basis of the maximum r
tax which has the characteristics of stamp duty charged on the acquisition of building l
s to fix the amount of the specific excise duty levied on the cigarettes under common ru
the effect of the increase in the rates of duty on spirits on 7 September 1977 by law no
y on beer - export refund - countervailing duty on imports . Case c - 152 / 89 . INDEX +
ning the application of the anti - dumping duty on ball - bearings and tapered roller be
to prove that the adjustment of the excise duty on beer leads to over - taxation of impo
rned with the imposition of anti - dumping duty on products assembled or produced in the
Belgo - Luxembourg Economic Union , excise duty on beer is levied in Belgium and Luxembo
tional measures introducing a differential duty on coal imported from the open market in
permit the Member States to impose capital duty on an interest - free loan granted by a
e having an effect equivalent to a customs duty on exports , as prohibited in trade betw
to exemption from turnover tax and excise duty on imports in international travel Havin
addition to the bound duty , an additional duty on sugar , corresponding to the charge b
to footnote (a) concerning an additional duty on sugar . This footnote provides that "
prices . 4 . Where necessary , the excise duty on cigarettes may include a minimum tax
ant whether the charge is in the form of a duty or tax or in the form of an equalization
an actual increase of the rate of customs duty or from a rearrangement of the tariff re

Pre-nominal and post-nominal modifications prevail in N-1 and N+1 positions, but its collocates are different if compared to "tax":

dumping (716)
customs (617)
excise (598)
definitive (308)
free (296)
imports (285)
rate (259)
provisional (257)
subject (160)
products (141)

Terms like "dumping" or "customs" do not collocate with "tax", nor does "turnover" collocate with "duty". Through the term "income tax", direct taxes are exemplified whereas through "excise duties" indirect taxes are exemplified. Duty is a tax levied on

commodities, transactions or estates rather than on persons. It is an indirect tax. On closer inspection of the collocates of "tax" and "duty", we see that in the first group of examples, where "tax" occurs, reference is primarily made to direct taxation, while in the second group of examples, where "duty" occurs, reference is primarily made to indirect taxation. In English a primary distinction is drawn between direct and indirect taxation. In this distinction, a deviant example can be found in the occurrence of "VAT" and "value-added tax", a tax paid on the supply of all goods and services in the U.K., introduced in 1973 to harmonize the British tax system with that of the other European Community countries. The occurrence may be explained by the general character acquired by the tax and by the superordinate value that the term "tax" holds.

5.3. A cross-linguistic comparison

If we consider the data of the Italian subcorpus we find significant similarities and differences in the translation equivalents.

As to the first meaning of "tax", for instance, it will be observed that a class shift is implied as the adjective "fiscale" (1,696) appears to be its translation equivalent in Italian, collocating with such words as "sistema", "carico", "franchigia", "deposito", "esenzione", "evasione", etc.. As we have seen, this may be traced back to the different composition of noun groups in English and Italian:

no . oppure il diritto a tale agevolazione fiscale spetti solo nel caso in cui l ' alcoo
 ia i reclami rivolti all ' amministrazione fiscale , sia i ricorsi giurisdizionali . 12
 iudice d ' appello , l ' amministrazione fiscale ha riconsiderato la sua posizione . e
 ulio vacanze e sottraendone l ' anticipo fiscale e gli oneri sociali a carico del lavo
 venir assimilati ad essa sotto l ' aspetto fiscale , e di respingere il ricorso per il r
 seconda dei casi , nella stessa categoria fiscale , doganale o statistica . b) il 2)
 particolare all ' efficacia del controllo fiscale o , ai sensi dell ' art . 36 del trat
 stinguere quindi interamente il suo debito fiscale , presentando pero le sue rimostranze
 , di conseguenza , al sorgere di un debito fiscale in fatto d ' imposta sulla cifra d ' r
 o membro in cui e' autorizzato il deposito fiscale ; >> . 4) all ' articolo 14 e' aggiu
 o del diritto delle societa' e del diritto fiscale . altre legislazioni riconoscono alle
 uto che il cantisani , nella dichiarazione fiscale dei redditi per il 1977 , aveva dichi
 ffermato che il divieto di discriminazione fiscale di cui all ' art . 95 del trattato ce
 usare autoveicoli importati in franchigia fiscale sarebbe un mezzo necessario , in quan
 evitare il rischio di evasione o di frode fiscale . in particolare , non e' provato che
 to alla " tax evasion " , cioe' alla frode fiscale . 30 e' opportuno osservare che , dal
 igidamente il principio dell ' imposizione fiscale nello stato membro destinatario , il
 iari . secondo le disposizioni della legge fiscale , il mutuatario puo' dedurre dall ' i
 ' istituire tributi che non abbiano natura fiscale , ma siano istituiti specificamente p
 ione contraria al principio di neutralita' fiscale inerente al sistema comune di imposta
 ida in modo apprezzabile sul futuro onere fiscale , devono essere fornite indicazioni i
 nte dev ' essere raffrontato con l ' onere fiscale pu' ridotto effettivamente sopportato
 ro non e' sottoposto ad alcun provvedimento fiscale o di effetto equivalente che nella su
 - 1 , lett . b) , del codice di procedura fiscale (livre des procedures fiscales) dec
 a quale era volta a disciplinare il regime fiscale in modo tale da farlo rimanere , in r
 sulla questione relativ... al diverso regime fiscale per le autovetture usate importate e
 protezionistico di un determinato sistema fiscale nazionale ; orbene , risulta che , no
 destinati all ' esportazione in un sistema fiscale volto a finanziare il controllo dei m
 opporre i vini importati ad un sovraccarico fiscale atto a proteggere la birra di produzi
 lare implicante un determinato trattamento fiscale , l ' analogo prodotto importato , ai

As far as meanings 2 and 3 are concerned, a parallel can be drawn between the occurrences of "tax" in the English subcorpus and of "imposta" in the Italian one. In a high percentage of cases, "tax" finds its counterpart in "imposta". "Imposta" like "tax" is used as a superordinate, but if we consider the collocates of "imposta", we notice relevant differences in the collocations of the two terms.

Let us have a quick scan through the concordance of "imposta" (4,209):

ria , la legge olandese relativa all ' imposta sulla cifra d ' affari ha previsto mod
ciplina esauriente delle fanchigie dall ' imposta sull ' entrata e dai diritti d ' acc
one il bene e , di fatto , gravato dall ' imposta soltanto in base al valore aggiunto in
e il cliente e' registrato ai fini dell ' imposta sul valore aggiunto ; - l ' opera fabb
upero dei prelievi rispetto a crediti d ' imposta analoghi ai quali gli stati membri ric
che prescrive il metodo di calcolo dell ' imposta di conguaglio da applicare nei loro co
tti agricoli . la parte " mobile " dell ' imposta contemplata dall ' articolo 10 sopra c
l quale egli e' registrato ai fini dell ' imposta sul valore aggiunto e destinati alla p
72 , relativa alle imposte diverse dall ' imposta sulla cifra d ' affari che gravano sul
erazione per determinare l ' aliquota d ' imposta applicabile ai redditi della moglie di
di assoggettare detta retribuzione all ' imposta nazionale sul reddito . di conseguenza
le . la natura protezionistica di quest ' imposta e accentuata dal fatto che essa ammont
el procedimento c - 353 / 90 " 1) se l ' imposta sul consumo delle banane fresche , int
gine , in via di principio , a debiti d ' imposta sulla cifra d ' affari all ' importazi
di merci cedute da privati , qualora un ' imposta del genere non venga riscossa sulla ce
a direttiva osti alla riscossione di un ' imposta speciale sugli spettacoli e sugli intr
a sia la struttura che le aliquote dell ' imposta stessa ; considerando che il mantene
colo 2 1 . le operazioni sottoposte all ' imposta sui conferimenti sono tassabili uncam
societa' di capitali . articolo 5 1 . l ' imposta e' liquidata : a) nel caso della cost
embri hanno la facolta' di riscuotere l ' imposta soltanto man mano che i conferimenti s
, e , di conseguenza , gli sgravi dell ' imposta sulla cifra d ' affari e delle altre i
ma si e pronunciata per il rinvio dell ' imposta sul valore aggiunto in italia al 1 gen
azi doganali dalla base di calcolo dell ' imposta proporzionale riscossa sulle sigarette
to una deduzione totale o parziale dell ' imposta sul valore aggiunto . tuttavia , i pre
neratore dell ' imposta si verifica e l ' imposta diventa esigibile all ' atto della ces
a legge tributaria ; l ' incidenza dell ' imposta controversa sui redditi comuni e incon
acente parte del sistema nazionale dell ' imposta sull ' entrata . * / 667 j0007 / * . u
imposta sull ' entrata col sistema dell ' imposta cumulativa a cascata e , in secondo lu
istituto , e calcolata in ragione dell ' imposta sul reddito pagata dai genitori , con
al fine di determinare l ' aliquota della imposta dovuta su altri redditi non esenti nel

A further difference is to be pointed out. Position N-1 is generally occupied by a definite article and "imposta" is generally modified on the right. N+1 and N+2 positions are generally occupied by post-modification.

In English data we find:

pre-modification + noun

while in Italian data we have:

[determiner] + noun + post-modification

The different structure of the noun group plays a role which cannot be overlooked and which will be the object of further analysis.

It is interesting at this point to compare "duty" with "tassa" as we might expect it to be its equivalent. But we see that the occurrences of "tassa" are definitely lower as the term occurs in 1,398 citations. Some of them, selected automatically, are reproduced here:

dei dazi doganali . per stabilire se una
, in determinati casi , l ' esonero dalla
mobilistica b) addizionale del 5 % sulla
e delle caratteristiche essenziali di una
tati direttamente da paesi terzi , di una
riscossione , da parte della pbc , di una
ro l ' italia , in merito alla stessa '
7 maggio 1987 , dichiara : un sistema di
nale propriamente detto , costituisce una
ere la seconda questione nel senso che la
gli stessi criteri , puo' costituire una
i tratta di un onere unico , denominato '
tassa abbia effetto equivalente a quello di un
tassa all ' esportazione per le patate (gu l
tassa automobilistica - lussemburgo : taxe sur
tassa del genere . a norma dell ' articolo 11
tassa destinata a scopi previdenziali . 3 le q
tassa destinata a sovvenzionare lo smercio al
tassa di sbarco ' , un ricorso per inadempim
tassa di circolazione che , mediante l ' istit
tassa di effetto equivalente ai sensi degli ar
tassa di compensazione riscossa sui vini greci
tassa di effetto equivalente ad un dazio dogan
tassa di presentazione in dogana ' . le due

fissa versato e l'importo massimo della tassa differenziale sulle autovetture di fabbricazione, di un contributo, di una tassa d'iscrizione o di un "minerval", con la legge 16 gennaio 1985, n. 13, sulla tassa d'immatricolazione degli autoveicoli e successive modifiche di detta legge, di una tassa d'immatricolazione sulle automobili e interpretato nel senso che esso colpisce la tassa postale per la presentazione in dogana doganata al trattato cee, comprenda anche la tassa scolastica percepita in base alla legge la controversia verte sul pagamento della tassa scolastica richiesta ad un dipendente di veicoli pesanti e riduzione parallela di una tassa sugli autoveicoli versata dai vettori na

If we consider the collocates, we find that the word "tassa" is modified by adjectives, such as "automobilistica", "postale", "scolastica" and by noun groups such as "di circolazione", "d'immatricolazione", "d'iscrizione". The reference to direct and indirect taxation is not made in the distinction drawn in Italian between "imposta" and "tassa". Different conceptual categories are applied in the two languages. "Tassa automobilistica", which finds its equivalents in the corpus data both in "vehicle tax" and in "vehicle duty", is something paid for a consideration of value. A payment is due in return for services.

An outstanding feature of Italian tax law is the distinction made with regard to contributions levied on a person with or without regard to personal services or advantages conferred on that person by law. The word "tassa" occurs when the payment is meant as a counterpart of personal or general services.

6. Conclusion

The analysis should be extended to include other terms such as "charge", "rate", and "fee". Work is in progress. Even limiting our consideration to the terms under scrutiny, we can say that through the analysis of the collocates, the legal framework of the tax law emerges in its main outlines showing, through the collocates, relevant differences between the systems of civil law and common law.

On the one hand, corpus evidence suggests that collocation plays a fundamental role in the definition of words. On the other, this shows that, in a number of cases, the origins of linguistic differences are to be sought in institutional and historical traditions of different countries as extrinsic forces may play a part in the semantic determination of the words under scrutiny. This raises a number of questions, but as a partial conclusion of our study we can say that by making such empirical information available corpus linguistics may provide the tools for semantic analysis. As the development of special corpora continues and provides a more adequate database upon which to address questions, they ought to play an increasingly important role in linguistic description. We think that more research should be conducted in this direction.

REFERENCES

- Aijmer, K. & Altenberg, B. (eds.), 1991, *English Corpus Linguistics*, London-New York, Longman.
- Baker, M., Francis, G. & Tognini-Bonelli, E. (eds.), 1993, *Text and Technology: in honour of John Sinclair*, Amsterdam, Benjamins.
- Atkins, S., Clear, J. & Ostler, N., 1992, "Corpus design criteria" in *Literary and Linguistic Computing*, 7, 1, Oxford, Oxford University Press, 1-16.
- Biber, D., 1983, "Representativeness in corpus design" in *Literary and Linguistic Computing*, 8,4, Oxford, Oxford University Press, 243-57.
- Hart, H.L.A., 1953, *Definition and Theory in Jurisprudence*, Oxford, Clarendon Press.
- Mason, O, 1996, *Corpus access software: The CUE system*, TEXT Technology, 6, 4, 257-266.
- Reichard, K. & Johnson, E.F., 1996, *Using XForms*, Unix Review, 84.
- Rossini Favretti, R, 1993, "Estate e tenure come espressione del concetto di proprietà feudale" in *Aspects of English and Italian Lexicology and Lexicography*, 244-53, Hart, D. (ed.). Roma, LIS.
- Rossini Favretti, R., 1999, "Scientific discourse: intertextual and intercultural practices" in Rossini Favretti, R., Sandri, G. & Scazzieri R. (eds.), *Incommensurability and Translation*, Cheltenham, Edward Elgar.
- Rossini Favretti, R. "Using multilingual parallel corpora for the analysis of legal language: the Bononia Legal Corpus", in Teubert, W., Tognini Bonelli E. & Volz, N. (eds.), *Proceedings of the Third European Seminar, Translation Equivalence*, The TELRI Association e.V., Institut für deutsche Sprache, Mannheim, The Tuscan Word Centre, 57-68.
- Sinclair, J.M., 1986, "First throw away your evidence" in *The English Reference Grammar*, 56-65, Leitner, G. (ed.), Tübingen, Niemeyer.
- Sinclair, J.M., 1987, *Looking up*, London and Glasgow, Collins.
- Sinclair, J.M., 1991, *Corpus, Concordance, Collocation*, Oxford, Oxford University Press.
- Sinclair, J.M., 1995, "Corpus typology. A framework for classification" in Melchers G. & Warren, B. (eds.), *Studies in Anglistics*, Stockholm, Almquist and Wiksell International, 17-34.
- Sinclair, J.M., 1996, " Multilingual databases. An international project in multilingual lexicography", in *International Journal of Lexicography*, 9,3, 179-96.
- Stubbs, M, 1995, "Collocations and semantic profiles", in *Functions of Language*. 2, 1, 23-55.
- Svartvik, J.(ed.),1992, *Directions in Corpus Linguistics*, Berlin-New York, Mouton de Gruyter.
- Teubert, W., 1996, "Comparable or parallel corpora?" in *International Journal of Lexicography*, 9, 3, 238-64.
- Thomas, J.& Short, M. (eds.), 1996, *Using Corpora for Language Research*, London-New York, Longman.
- Zhao, T. C. & Overmars, M., 1995, *Forms Library. A graphical user interface toolkit for X*, <http://bragg.phys.uwm.edu/xforms>.