

Ridings, D. (1996): Text representation in PAROLE. Parole MLAP 63-386, Work package 4.1.3

Sperberg-McQueen, C.M./L. Burnard (eds.) (1994): Guidelines for Electronic Text Encoding and Interchange. Chicago and Oxford.

R. Rossini Favretti, Bologna

Using Multilingual Parallel Corpora For The Analysis of Legal Language: The Bononia Legal Corpus

Abstract

This paper intends to describe our experience in building multilingual comparable law corpora here at the University of Bologna. It is the first report of an ongoing corpus-driven research project named Bononia Legal Corpus, or BOLC, which is involved in investigating legal language in English and in Italian.

In the construction of the corpus, two subcorpora have been envisaged: one English, the other Italian. According to our plans, carrying out the project will take three years: 1997-1999. As to the dimension of the corpus, we set 10 million words as the smallest target for each component.

Before the first sample, a preliminary pilot corpus (6 mw for each subcorpus) was stored.

KEYWORDS: corpus linguistics, multilingual parallel corpora, legal language.

1. Introduction

This paper intends to describe our experience in building multilingual comparable law corpora here at the University of Bologna. It is the first report of an ongoing corpus-driven research project named Bononia Legal Corpus, or BOLC, which is involved in investigating legal language in English and in Italian.

BOLC is an interdisciplinary project. It is based on the cooperation of experts in linguistics, computer programming,¹ and law². The presence of legal experts in the team has proved particularly important in drawing up specimen bibliographies, in the selection and classification of texts (on cultural evidence), in transmuting parallel corpora into comparable corpora, and in the selection of key concepts and key words. The presence of John Sinclair as a consultant has also been of critical importance.

¹ As computer experts I wish to mention F. Tamburini and E. Martelli for their valuable work in formatting corpus documents and writing the software which is used in our research.

² As a legal expert I wish to mention in particular Prof. A. Di Pietro.

In the construction of the corpus, two subcorpora have been envisaged: one English, the other Italian. According to our plans, carrying out the project will take three years: 1997-1999. As to the dimension of the corpus, we set 10 million words as the smallest target for each component, that is:

first sample	10 m. w.	English	10 m.w.	Italian
subsequent	" "	" "	" "	" "
each sample	" "	" "	" "	" "

The samples will be restricted to readily available material; 90% or more is probably already in electronic form.

Before the first sample, a preliminary pilot corpus (6 mw for each subcorpus) was stored.

2. A legal corpus

Why a legal corpus? For the relevance that collocation has in legal texts. It has been argued that when we think of law, we usually think first of words. This is an important point since most juridical words do not have a direct connection with counterparts in the actual world. They cannot be defined in terms of ordinary factual counterparts; concordance and collocation, therefore, play a fundamental role in the definition of meaning.

English and Italian legal texts are representative of two different legal systems and of the differences existing between the common law system developed in England and the civil law system, based on the Roman law, developed in Italy.

A further point is to be made as to the representativeness of the corpus design. In the last few years, the situation has been made more complex with the development of European Community law, that body of law created by and consequent to the association of European countries. European Community law not only applies within the Community as an organization, but it also binds member states together, and creates rights and duties for legal persons therein.

In our project, European legislation is seen as a reference for the contacts taking place there among the different legal systems, and for the transnational approach which is implied in the coexistence and cooperation of different nationalities. In the building up of our corpus, we chose to refer to European legislation, in a preliminary pilot corpus, in order to create a conceptual framework to be used as a first-level reference. We chose to refer to the secondary Community legislation even if reference can also be made to primary legislation, that is to the Treaties that established the communities.

Secondary legislation consists of regulations, decisions, directives, recommendations and opinions made or given by the Commission. In particular, our focus was on directives since they can be implemented by domestic legislation and can produce direct legal effects in member states. At a further stage, directives make possible the identification and selection of legal texts which have been enacted "on specific issues," in different European countries. They make it possible to transmute parallel corpora into comparable corpora.

Data are now limited to English and Italian. An extension which will include French and Spanish is envisaged.

A short description of the documents. A directive usually begins with an extensive set of header information, indicating the date, the reference number and its coding conventions. The text follows. After the proper text there is possibly some trailer information. The text data have been divided from the header and the trailer data. Two documents have been distinguished.

The first document (data file) contains the legal text and the reference number (CELEX).

The second document (reference file) contains all other data that could be useful in performing law-targeted rather than language-targeted searches.

During the first phase, while building the pilot corpus, the corpus was not tagged in order to simplify the entire building process. In the future, the tagging process could be performed with some differences between subcorpora, in consideration of the specific requirements presented by Italian, as, for example with accents and apostrophes.

3. A case study: *tax and duty*

In the analysis of the pilot corpus some important methodological aspects emerged. I will comment briefly on one of the most significant collocations³ with regard to the problems presented in interlinguistic comparison. As a case study, I will consider the English word: *tax*. It refers to a situation which can be considered common both to the U.K. and to Italy, and can be assumed to apply to other European countries as well, and whose presence can be accounted for with the extension of our corpus. In all countries, taxes are levied on income and expenditure by central and local governments, but different

³ "Collocation is the occurrence of two or more words within a short space of each other in a text. The usual measure of proximity is maximum of four words intervening. Collocations may be dramatic and interesting because unexpected, or they can be important in the lexical structure of the language because of being frequently repeated." J. SINCLAIR, 1991, p. 170.

categories are employed in how taxes are defined. It was our hypothesis that the categories would emerge from the analysis of the data.

Let us consider the following concordance of the word *tax*, selected automatically from our corpus, where there are 1204 citations altogether:

Article 7 1. The term ' withholding tax ' as used in this Directive shall n
to maintain temporarily a withholding tax , HAS ADOPTED THIS DIRECTIVE : Articl
the calculation of a toll , tariff , tax , bonus , penalty , remuneration , in
exemption from or refund of turnover tax , excise duty or any other consumptio
by private individuals from turnover tax , excise duty and other consumption t
ing out of ordinary activities , after tax , for the last financial year . (g)
cified which persons are liable to pay tax , in particular as regards services s
, both net and inclusive of duty and tax , whether published or not , for a pr
ads of state Article 60 Exemption from tax , within the limits and under the con
of the tax base , calculation of the tax , chargeability and monitoring of the
es and saddle - horses from : Turnover tax , excise duties and any other consump
subject to prior payment of value added tax , at the rate applying on the date of
basis of the standard rate of turnover tax . Where that special rate covers one
into account exclusive of value added tax : 1 . The value of the total final pr
on measures with regard to value added tax THE COUNCIL OF THE EUROPEAN COMMUNITI
ries shall be exempt from the turnover tax and excise duty levied on imports if
lating to the rules governing turnover tax and excise duty applicable in interna
pe of the common system of value added tax as established by Council Directive 7
dispatched , make a declaration to the tax authorities of the Member States of d
chargeable event and the date when the tax becomes chargeable with those laid do
icle 3 For purposes of calculating the tax burdens on a product at the final sta
banks) Death duties Tax information , Tax deposits . Other capital transactions
esignated as the person liable for the tax due in accordance with Article 28 c ()
Member States will make provisions on tax exemptions and duty - free allowances
ber State which issued the value added tax identification number under which the
basis of assessment of the value added tax in order to apply the scheme of own r
fy the Community standards by means of tax incentives ; whereas this implies tha
States may adopt arrangements whereby tax is payable by another person . Inter
untries in such a way that a saving in tax may result in one or the other Member
his business , where the value added tax on the goods in question or the compo

On inspecting the concordances of *tax*, we can find some relevant collocations. *Tax*, in the singular, presents a high occurrence in our pilot corpus. In all instances nouns dominate the pattern at N-1 position both on the right and the left. *Tax* tends to occur either followed or preceded by a noun or a noun group. It occurs both i) as a modifier and ii) as a headword.

In a particularly high number of examples (406), it occurs as a modifier in a noun group. If we consider the top ten we have:

- tax authorities (35)
- tax exemption (30)
- tax system (30)
- tax burden (29)
- tax relief (26)
- tax warehouse (18)
- tax evasion (13)
- tax rules (12)
- tax representative (11)
- tax remission (9)

In a number of examples the word *tax* is associated with prenominal or post-nominal modification.

The first position to the left of *tax* is occupied:

-by a noun in 102 instances, which is roughly 9%:

- turnover tax (68)
- income tax (10)
- consumption tax (6)
- export tax
- vehicle tax

- by a noun-group in 366 cases:

- value added tax (264)

-by an -ing form in 23 cases:

- a withholding tax (11)

In a number of cases (94), *tax* is qualified by a *prepositional phrase*:

- tax on exports/export/exportation (14)
- tax on profit or loss (10)
- tax on services (9)
- tax on sales (1)

Through postnominal qualifications, the objects and services upon which taxes can be imposed are generally specified. As we can see in the examples, specification can take place both in postmodification and in premodification: *tax on export* or *export tax*.

It is interesting to observe that adjectives occur as modifiers in a limited number of instances, 19 examples out of 1204, which is lower than 2 %.

In the examples where the word *tax* is not associated with modification or qualification, the first position on the left is occupied:

-by a preposition:

- after tax (18)
- before tax (3)
- subject to tax (11)
- liable to tax
- the amount of tax (10)

ration of a product subject to excise duty shall mean the entry of that product into the territory of a Member State for the purposes of export or import, or for the purposes of charging capital duty, except in circumstances duly justified by the quantities set out in the Annex, failing which the quantities set out in the Annex shall be admitted free of duty in accordance with the national law of the Member State which replace an excise duty by a specific excise duty and the specific excise duty which is to be charged only on transaction duty may not be less than 5% or more than 10% of the value of the goods, on the basis of the most appropriate basis for levying duty shall be established by reference to the gravity bands of more than one duty Whereas a fair and effective application of the duty with a minimum or a maximum duty base without having to resort to the inwards

Excise or customs does not collocate with *tax*, nor does *turnover* collocate with *duty*. On closer inspection of the collocates *duty* and *tax*, we see that in the first group of examples, the noun groups, where *tax* and *taxes* occur, reference is made to direct taxation, while in the second group of instances, where *duty* and *duties* occur, reference is made to indirect taxation.

Through the terms, *turnover tax/taxes*, direct taxes are exemplified; through *excise duties*, indirect taxes are exemplified. In considering the English corpus data we find that *customs/excise duties* are levied on petrol, wines, tobacco, etc. *Duty* is a tax levied on commodities, transactions or estates rather than on persons. It is an indirect tax. 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, 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 E.C. countries. The occurrence may be explained by the general character acquired by the tax, and by the superordinate value that the word *tax* holds.

4. *Imposta and tassa in the Italian corpus*

If we consider the data of the Italian corpus, we find significant similarities and dissimilarities.

As to the first meaning of *tax*, a class shift is implied: the adjective *fiscale*, as we can observe going through the Italian corpus, is its translational equivalent

lente in Italian, collocating with such words as *sistema, carico, franchigia, deposito, esenzione, evasione, etc.*:

talune forme di frode o di evasione fiscale, HA ADOTTATO LA PRESENTE DIRETTIVA del progresso dell'armonizzazione fiscale, non consente ancora in Danimarca cora ammortizzate dal punto di vista fiscale, essi estendono il beneficio di taluna LL' APPLICAZIONE DELLA LEGISLAZIONE FISCALE, NELL' ALLEGATO OCCORRE INDICARE olli da parte dell'amministrazione fiscale. b) Ogni soggetto passivo deve te di previdenza sociale, - sul regime fiscale. 2. Gestione commerciale e finanz fiscale. Articolo 12 L' apertura e il fun lta, sono effettuati in un deposito fiscale; » . 4) All' articolo 14 è agglu bro in cui è autorizzato in un deposito fiscale; che le esenzioni all' importazio gli obiettivi dell'armonizzazione fiscale; che le esenzioni all' importazio RIPRESE PER IL CALCOLO DELL' ONERE FISCALE ANTERIORE, CONFORMEMENTE ALLE DISP OTATI . 2.5 . INDICAZIONE DEL REGIME FISCALE CONCERNENTE QUALSIASI IMPOSTA E TAS AGLI ESERCIZI PRECEDENTI E L' ONERE FISCALE GIÀ PAGATO O DA PAGARE PER TALI ESE COMMISSIONE IL CALCOLO DELL' ONERE FISCALE MEDIO, STABILITO IN CONFORMITÀ DEG RATO IN UNA FASE QUALSIASI L' ONERE FISCALE NELLE FASI ANTERIORI NON È STATO CA piena applicazione della franchigia fiscale accordata ai viaggiatori provenienti ; c) regime sospensivo: il regime fiscale applicabile alla fabbricazione, al opportuno garantire un'incidenza fiscale armonizzata per tutti i prodotti ap erando che le disposizioni di natura fiscale che si frappongono attualmente alla distorsioni; che soltanto un regime fiscale comune può pertanto costituire una fiscale dei redditi, gli organismi finanzia di funzionari dell'amministrazione fiscale dell'altro Stato membro. Le moda dei prodotti. Questo rappresentante fiscale deve essere stabilito nello Stato m funzionari dell'amministrazione fiscale di uno Stato membro sul territorio ne, con l'applicazione del regime fiscale e di altre forme di contributo prev dei sistemi fiscali sulla politica fiscale e di bilancio degli Stati membri no del consolidamento tra l'onore fiscale imputabile all'esercizio in corso ese abbia nominato un rappresentante fiscale in detto paese. Gli Stati membri p le sigarette può comportare un onere fiscale minimo il cui massimale è determina me obiettivo la frode o l'evasione fiscale o ha come conseguenza che una socie delle aliquote per motivi di equità fiscale o di ordine sociale ovvero per perm

This may be traced back to the different formation of noun groups in English and in Italian. In English a noun can be modified both by an adjective and by a noun. Most noun groups consist of two or more nouns – *tax law, etc.*. In Italian, they predominantly consist of a noun either preceded or followed by one or more adjectives – *diritto fiscale, etc.*. Prenominal modification prevails in English. Postnominal modification prevails in Italian. This can have an important bearing on our analysis of right and left collocates

As far as meanings 2 and 3 are concerned, a parallel can be drawn between the occurrences of *tax/taxes* in the English corpus and *imposta/imposte* in the Italian. In a high percentage of cases, *tax* finds a counterpart in *imposta* or *imposte*.

Imposta, like *tax*, is used as a superordinate and is used in the singular and in the plural with a number of nouns as collocates.

However, if we consider the collocates of *imposta*, we notice relevant differences in the collocation.

erisce la domanda di rimborso dell'imposta (ad esempio: partecipazione all' tato designato quale debitore dell'imposta, in conformità dell' articolo 28 re sia tenuto in solido a pagare l'imposta; » . 20) All' articolo 28 novies TRODUZIONE DEL SISTEMA COMUNE DELL' IMPOSTA SUL VALORE AGGIUNTO; CONSIDERANDO odotti a causa dell' alto livello d' imposta al quale sono attualmente assoggett della loro produzione, esclusa l' imposta che colpisce la vendita da parte de on sono autorizzati a dedurre dall' imposta che devono versare l' imposta sul può procedere alla deduzione dell' imposta che avesse eventualmente gravato su 26 bis, parte B, paragrafo 3 e l' imposta che si considera inclusa in detto p ro, una riduzione od un esonero d' imposta che dovrebbe comportare per esso un el Regno Unito, o a qualsiasi altra imposta che venga a sostituire una delle im

d'antiquariato assoggettate all'imposta conformemente alla lettera a) sono corso delle tappe precedenti, l'imposta di consumo è in linea di massima ri affari, accise o qualsiasi altra imposta di consumo. Per l'applicazione d ato e si stabilisce l'aliquota d'imposta di base applicabile per gli importi etti passivi e sono debentrici dell'imposta dovuta a titolo di acquisti intraco curare l'esatta riscossione dell'imposta e ad evitare le frodi, a condizio beneficenza della franchigia dall'imposta prevista all'articolo 24 e della proporzionale e l'importo dell'imposta specifica devono essere uguali per ari, di accise o di qualunque altra imposta sui consumi. Per l'applicazion che sono già stati sottoposti all'imposta sui conferimenti; - l'ammontare di dazi doganali, e concernente l'imposta sul valore aggiunto (86/489/CE passivo, registrato ai fini dell'imposta sul valore aggiunto all'interno d ottenuti, per quanto concerne l'imposta sul valore aggiunto, sia lo stesso Stati membri, del sistema comune di imposta sul valore aggiunto, senza la cont particolare le seguenti: Italia: Imposta sul reddito delle persone fisiche I azione della base imponibile di un'imposta sul reddito, sui benefici o sulle il 1° gennaio 1995, in materia di imposta sul valore aggiunto IL CONSIGLIO DE relativà alle imposte diverse dall'imposta sulla cifra d'affari che gravano

A further observation is to be made. Position N-1 to the left of *imposta* is generally occupied by definite article and the noun is generally modified on the right. N-1 and N-2 positions on the right are generally occupied by post-modification. The different uses of the article and the different structure of the noun group play an important role in a concordance which cannot be overlooked, and which will be the object of further analysis.

If we examine the corpora and focus on the instances where the enumerations occur, we find *taxes*, *duties* and *charges* mentioned as different types of contributions in English. The word *charges* tends to occur as a general, inclusive term, at the end of the enumeration; the most prominent collocates are *duty* and *tax*. It is interesting to compare *duty* with *tassa*, which one might expect to be its equivalent. But we see that the occurrences of *tassa* are definitely lower. There are 31 citations of *tassa*; every third one is reproduced here:

74, relativo alla riscossione di una tassa all'esportazione di taluni prodott stica, b) addizionale del 5% sulla tassa automobilistica; - Lussemburgo: ta erale, qualsiasi imposta, diritto e tassa che non abbia il carattere di impost ti sia stata pagata in anticipo e la tassa complessiva esigibile sui prodotti a a condizione che essi prelevino una tassa di controllo di ecu 5 per 1 000 litri effetto di rendere l'importo della tassa inferiore alle aliquote minime previ diritti d'utenza in funzione della tassa nazionale sugli autoveicoli. Questo ché tuttavia tale imposta, diritto e tassa non dia luogo, negli scambi fra Sta ri si sostituiscono a qualsiasi altra tassa o contributo sanitario riscosso dall herm. (2) Il prezzo al netto della tassa si ottiene direttamente dalle tariffe

There are 122 citations of *tasse*. Every fourth one is reproduced here:

te, degli oneri legali obbligatori (tasse, imposte e diritti) nonché delle s limiti previsti da spacci esenti da tasse, nel quadro del traffico di viaggio le imposte, i dazi, i prelievi e le tasse, - le spese accessorie quali le com le imposte, i dazi, i prelievi e le tasse, le spese accessorie quali commissi mente esenti da dazi e tasse e dazi e tasse compresi, pubblicati o no, di un p ncessa una franchigia dalle imposte e tasse di cui all'articolo 1 in caso di i ncessa una franchigia dalle imposte e tasse di cui all'articolo 1, per una du revolmente aumentare il livello delle tasse attualmente applicate agli autoveico me dovute a titolo di dazi doganali, tasse d'effetto equivalente o prelievi a sostanze nella base di calcolo delle tasse di circolazione dei veicoli a motore icare l'importo dei dazi doganali, tasse di effetto equivalente, prelievi ag portazione» sia i dazi doganali e le tasse di effetto equivalente sia i preliev esenzione dai dazi doganali, dalle tasse di effetto equivalente e dai preliev nire determinati i dazi doganali, le tasse di effetto equivalente e i prelievi ncessa una franchigia dalle imposte e tasse di cui all'articolo 1 in caso di i tato membro che sostituisca una delle tasse di cui al paragrafo 1 con un'altra della tariffa doganale comune, delle tasse di effetto equivalente, dei preliev mo autorizzata, i dazi doganali, le tasse di effetto equivalente ed i prelievi portazione: sia i dazi doganali e le tasse di effetto equivalente, sia i preli importazione», i dazi doganali e le tasse di effetto equivalente, i prelievi

itti e tasse su certi servizi, quali tasse di atterraggio, di sbarco o di imba oste, i dazi, i prelievi e le altre tasse dovuti fuori del paese di importazio stati ugualmente istituiti prelievi, tasse e importi compensativi monetari all li Stati membri possono riscuotere le tasse elencate nell'allegato della prese oio mediante l'armonizzazione delle tasse o delle accise, ma che, in attesa la conoscenza della fiscalità e delle tasse parafiscali esistenti in ogni Stato relativà alla soppressione di alcune tasse postali per la presentazione in dog na somma non eccedente e diritti e le tasse richiesti ai propri cittadini per il vi al pagamento delle imposte e delle tasse secondo le disposizioni legali del p ssogettati ad aliquote ridotte delle tasse sugli autoveicoli; che, ai fini d

If we consider the collocates, we find that the word *tassa* is modified by adjectives, such as *automobilistica*, and noun phrases and prepositional phrases, such as *di circolazione*, *sulla circolazione*, *di atterraggio*, *di soggiorno*. The reference to direct and indirect taxation is not made in the distinction drawn in Italy between *imposta* and *tassa*. Different conceptual categories are applied in English and in Italian. *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.

In the plural form, the word *tasse* presents a more general value, which includes unspecified taxes and charges of many kinds, and finds its equivalent in *charges*:

Tasse di effetto equivalente charges having equivalent effect

By the words *tax*, *duty* and *tassa*, *imposta* different kinds of contributions are identified. Even limiting our analysis to positions N-1, N-2, N-3, we can discover relevant differences in the prominent collocates, and distinctions can be made between the different law systems. This brings to the fore another notable feature of our research. Legal language is deeply rooted in the institutional and historical traditions of a country. Meanings are bound to institutional practices. Further distinctions emerge and our analysis is to be extended to include not only *tax*, and *duty*, but also other words such as *charge*, *rate*, *fee*, etc., highlighting both their collocation and the equivalence that can be identified for translation. It should also include a closer examination of the cases in which difficulties may arise from the lack of syntactic correspondence between English compounds such as *tax-payer* or *chargeable event*, and Italian nouns and noun groups such as *contribuente* or *fatto generatore dell'imposta*. This is a further step, which can be made using corpora, and through practices which are not so different from the ones used in bringing forth and highlighting the equivalence between *tax* and *imposta* – an equivalence, it should be noted, which is not mentioned in any dictionary.

References

- Aijmer, K. and Altenberg, B. (eds.) 1991. *English Corpus Linguistics. Studies in Honour of Jan Svartvik*, London and New York, Longman.
- Baker, M, Francis, G. and Tognini-Bonelli, E. (eds.) 1993. *Text and Technology: in Honour of John Sinclair*, Amsterdam, Benjamins.
- Biber, D. 1994. "Representativeness in corpus design". In: *Current Issues in Computational Linguistics: in Honour of Don Walker*, Zampolli A., Calzolari N., Palmer M. (eds.), Giardini & Kluwer.
- Rossini Favretti, R. 1994. *Reading Legal Texts in English*, Bologna, Pitagora.
- Sinclair, J. 1991. *Corpus Concordance Collocation*, Oxford: Oxford University Press.
- Sinclair, J. 1996 (a). "Multilingual Databases. An International Project in Multilingual Lexicography", *International Journal of Lexicography*, 9, 3.
- Sinclair, J. 1996 (b). "The search for units of meaning", *Textus*, IX, 2.
- Svartvik, J. (ed.). 1992. *Directions in Corpus Linguistics*, Berlin-New York, Mouton de Gruyter.

Feng Zhiwei, Beijing

A study of translation equivalence in the Chinese-English text of Plato's *Republic*

Abstract

Based on the English-Chinese parallel corpus of Plato's *Republic*, we study the translation equivalence in the Chinese-English text of the corpus. Our study concentrates on the first chapter "Convention under Attack" and is focusing on three levels: paragraph level, sentence level and word level. The paper describes the agreement and disagreement between English and Chinese in these three levels. Finally an experiment of word alignment based on the corpora is introduced.

1. English-Chinese Parallel Corpus of Plato's *Republic*

Plato's *Republic* is a very famous and important philosophical monography. It is a sprawling work, and it also sprawled to China. In 1957, a Chinese translation from Greek was published, the translator is Wu Xianshu. Wu's translation was rather archaic and abstruse. In 1986, another Chinese translation version (also translated directly from Greek) was published, the translators are Guo Binhe and Zhang Zhuniing, both being professors of the history of philosophy. Guo/Zhang's version is easy and smooth, it is very popular in China and it has attracted a lot of Chinese readers. We decided to take Guo/Zhang's version as our study material. It was then manually put into electronic format as a corpus of Plato's *Republic* in Chinese translation. In 1996, Professor John Sinclair sent me a diskette containing a machine-readable corpus of the English translation of Plato's *Republic*, translated by Benjamin Jowett, and published in 1901 by Colonial Press. At the same time, Dr. Wolfgang Teubert sent me the book of this English version. With the help of Dr. Wolfgang Teubert and Professor John Sinclair, we set up a Chinese-English parallel corpus of Plato's *Republic*, where both versions were translated from the same language – Greek. The volume of the Chinese version is 443 136 bytes, and the volume of English version is 653 318 bytes.

Based on this parallel corpus, we started to study the translation equivalence in the Chinese-English text of Plato's *Republic*. Our study was concentrated to the first chapter "Convention under Attack". For the first chapter, the Chinese volume is 43 648 bytes, and the English volume is 67 200 bytes.