

## **Joined-up Terminology - The IATE system enters production**

**Sylvia Ball**  
SILD Division  
European Parliament  
ADG06C011,  
L-2929 Luxembourg

([sball@europarl.eu.int](mailto:sball@europarl.eu.int))

### ***Abstract***

The IATE<sup>1</sup> project for the development of a single terminology platform for the European Union's institutions and agencies began in 2000 and, although the system was originally scheduled to go live in 2001, it has only recently gone into full-scale user testing prior to the production phase. The paper describes the background to the project, the system implemented and the accompanying structures. It also discusses issues outstanding and future needs.

### ***Introduction***

The title of this paper is taken from the current (December 2002) online OED's first draft definition of "joined-up" in the figurative sense, viz.: "of thought, speech, etc.: coherent, organised, cogently or articulately developed". This aptly characterises the aspirations of the IATE project: to reorganise the terminology activities of the European Union (EU) institutions and agencies in a coherent manner, to eliminate the duplication of effort between institutions and consequent duplicate entries in the various terminology databases managed by them and to develop a single database for future activity using resources as rationally as possible, particularly with an eye to the enlargements of the EU scheduled to take place from May 2004 onwards.

It would be difficult to over-emphasise what an ambitious project this has proved to be, since it has meant bringing together the terminology databases of the three largest Community institutions, the European Commission (Eurodicautom), the European Parliament (Euterpe) and the Council of the EU (TIS), together with the databases more recently developed by other participants such as the Translation Centre (EuroTerms), the European Court of Auditors and the Court of Justice of the European Communities. Each had its own structure and philosophy, and it has been no easy matter to bring them together to make a coherent whole.

The paper will cover four main areas, the development of the project, in particular since November 2001 when it was last presented to an Aslib audience at T&C 23, the

---

<sup>1</sup> All abbreviations and acronyms are listed at the end of the paper with their explanation. For "IATE" in particular, see the section on the background to the project below.

functionality of the actual HTML implementation (this will be the core of the demonstration at T&C 25), the organisation of terminology activity in the production phase and, last but not least, future activity including the “joining-up” of legacy data.

## ***Background***

The IATE project was initiated in 1999 by the Translation Centre for the bodies of the European Union, hence its initial full title of the Inter Agency Terminology Exchange. The Centre, set up in 1994 to translate for the decentralised agencies of the EU, needed an online database to “provide adequate terminology facilities”<sup>2</sup> for its customers. In point of fact, the “Agency” aspect of the title was already outdated by late 1999 when the contract for development of the system was awarded to the Greek software firm Q&R and their linguistic sub-contractors the Danish Centre for Speech Technology (CST), since the large Community institutions mentioned above had already begun to take part. However, it was not until 2002 that the decision was finally taken to change the full title to “Interactive Terminology for Europe”. By then terminologists were all used to talking about IATE without being too concerned with what the acronym was supposed to stand for and we wanted to keep it. We also knew that we had to have an interactive system because that was what all our translator colleagues were clamouring for<sup>3</sup>, so the transition to “Interactive” was quite natural and appropriate.

Development proper began in early 2000, with July 2001 as the first, ambitious deadline for the system to be ready. Because of the complexity of the undertaking, it was in fact only a prototype that was ready for testing in August 2001. The prototype phase was followed by a number of pilot phases which gradually led to the development of the present system “combin[ing] functionality and user-friendliness”<sup>4</sup>. In the early stages of the project development was based on Oracle forms, which had numerous disadvantages. In June 2001 it was decided that the system had to be re-implemented in HTML and in April 2002 the developers presented what they called the “new navigational concept”, which is the basis of the system described here (minus one or two frills which professional terminologists felt were superfluous to their - or their translators colleagues’ - requirements). Keeping costs under control has been an important consideration. In an effort to reduce delays, a “workshop” method of development was adopted in September 2002, with one representative per contributing institution working directly with the developers to solve problems and improve the system's functionality and aesthetics.

## ***The functionality of the HTML implementation***

So much for the history of the project, but what did we actually end up with? As currently implemented, the IATE system allows users a full range of the functionality expected of any terminology database: consultation, data manipulation, and user management, as well as a few less usual additions, in our case mostly concerned with communication. When a project concerns such a large number of participants from different institutions, focused communication tools are a must.

---

<sup>2</sup> Macphail 2000, which is the main source for the early history of the project; see also Rummel/Ball 2001.

<sup>3</sup> Vidick/Defrise 1999

<sup>4</sup> Rummel/Ball 2001

The system is designed to be used by both institutional users, whether translators or terminologists, and the general public, with obviously a simpler interface for the latter class of users. Connecting to the public system brings up the homepage shown in figure 1. The homepage for institutional users is similar but, of course, password protected.

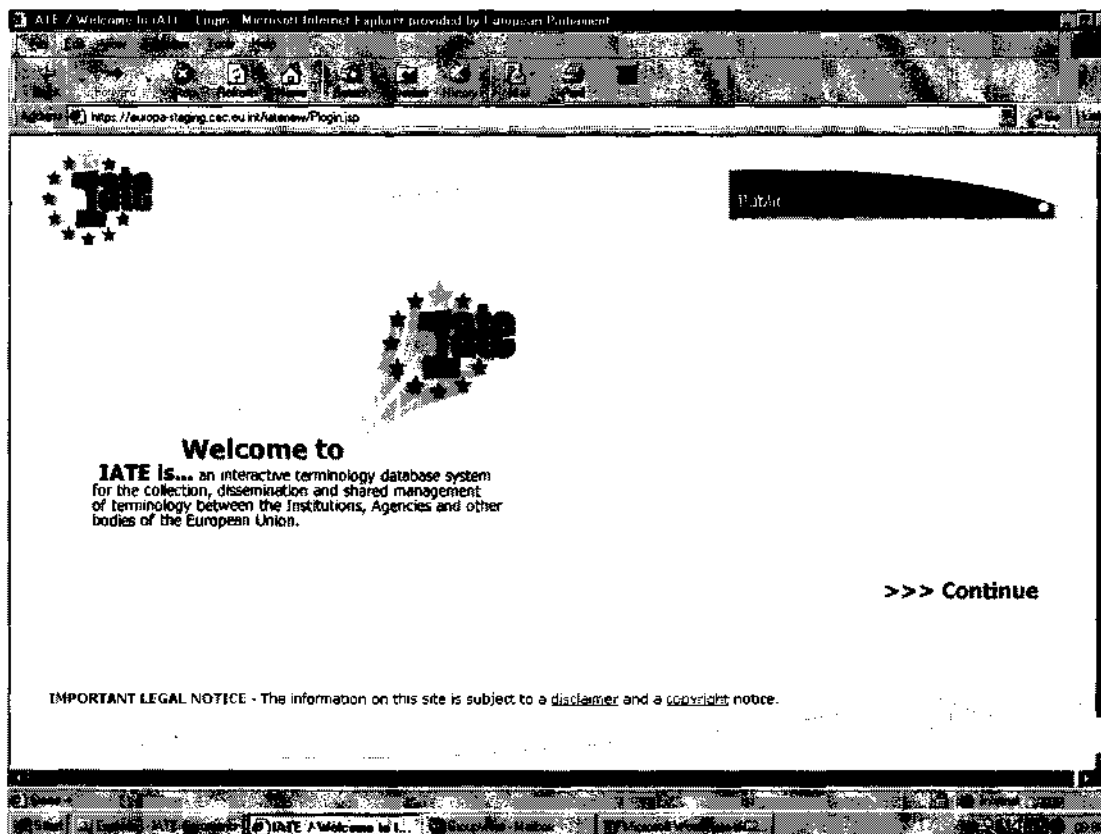


Fig. 1 The opening screen of the IATE public interface

For most translators, institutional or external, looking up problem terms as they translate will probably be the new system's primary function. It will simply take over from whatever resource they relied on before. Such a search could, theoretically, be as basic as looking for a particular character string, say "ABC", without specifying anything more<sup>5</sup>. Needless to say it is an approach which would give acceptable results only with a much better focused example, say "comitology", which returns only four hits even with no restriction as to language and the vast amount of data included in the system<sup>6</sup>. However, it seems reasonable to suppose that users who want to look up that particular term will at least specify that the language is English and that the search string is a term<sup>7</sup>.

<sup>5</sup> ABC with no restriction returns 213 hits, but you would only see the first 100, unless you at least selected an extended search (see Fig. 2, top right).

<sup>6</sup> Approximately 1.5 million concepts in all, in the data set of January 2003.

<sup>7</sup> Institutional translators will not need to set the conditions for each search since they can be part of the user's profile (see below in the comments on saving preferences).

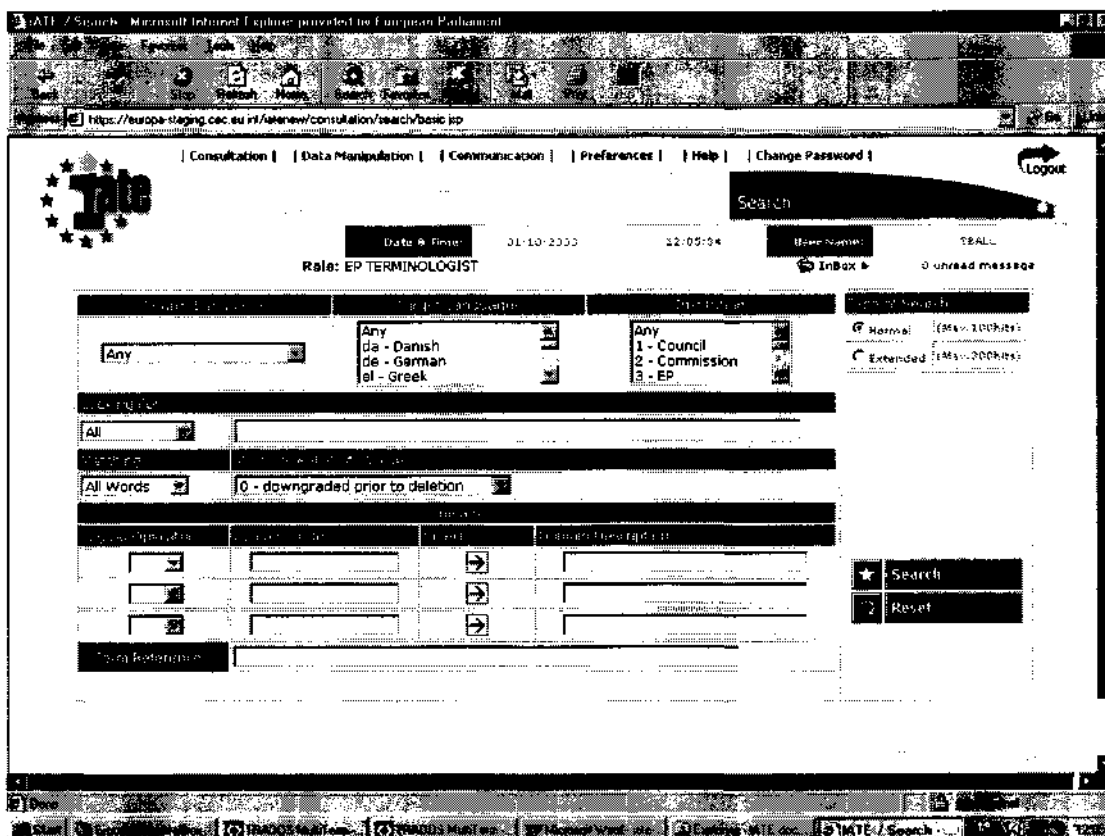


Fig. 2 Search screen

The layout of the search screen is in fact intended to lead the user to select at least the source language<sup>8</sup>, the type of string being looked for (term/phrase, term only, phrase only, abbreviation or formula) and the type of match required (from exact to fuzzy via “all words” and “any word”<sup>9</sup>), whether by profile or on the fly, since these elements are all to the left of the screen, before entering the search term followed by Enter or a click on the Search button to start the search. The selection criteria positioned above and below the search window (target languages, institution, subject domain<sup>10</sup>, reliability value and term reference) are intended to refine the search, where necessary (as in the case of ABC above, where selecting English as the source language only brings the number of hits down to 109, but adding French as the target gives 14, a much more manageable proposition, which a user is likely to feel able to evaluate).

The result of the search is displayed in a hitlist (figure 3) with 10 results per page. If no source language is specified, as in the search described for ABC, the hitlist terms are all shown in black, with hyperlinks (in blue) to a full display. When a source language is specified, as is suggested for “comitology”, the source language terms are shown in the hitlist in red. Clicking on a hyperlink brings up the full display (figure 4).

<sup>8</sup> In the list of source languages, the current EU languages are grouped together first by ISO code, followed by those of the 2004 Accession States.

<sup>9</sup> You would have to be fairly desperate for a solution to select “fuzzy” or “any word” from such a large data set.

<sup>10</sup> EuroVoc descriptors are used for domain classification; they are widely used for European documentation, available in all EU and accession languages, and maintained by a multilingual group.

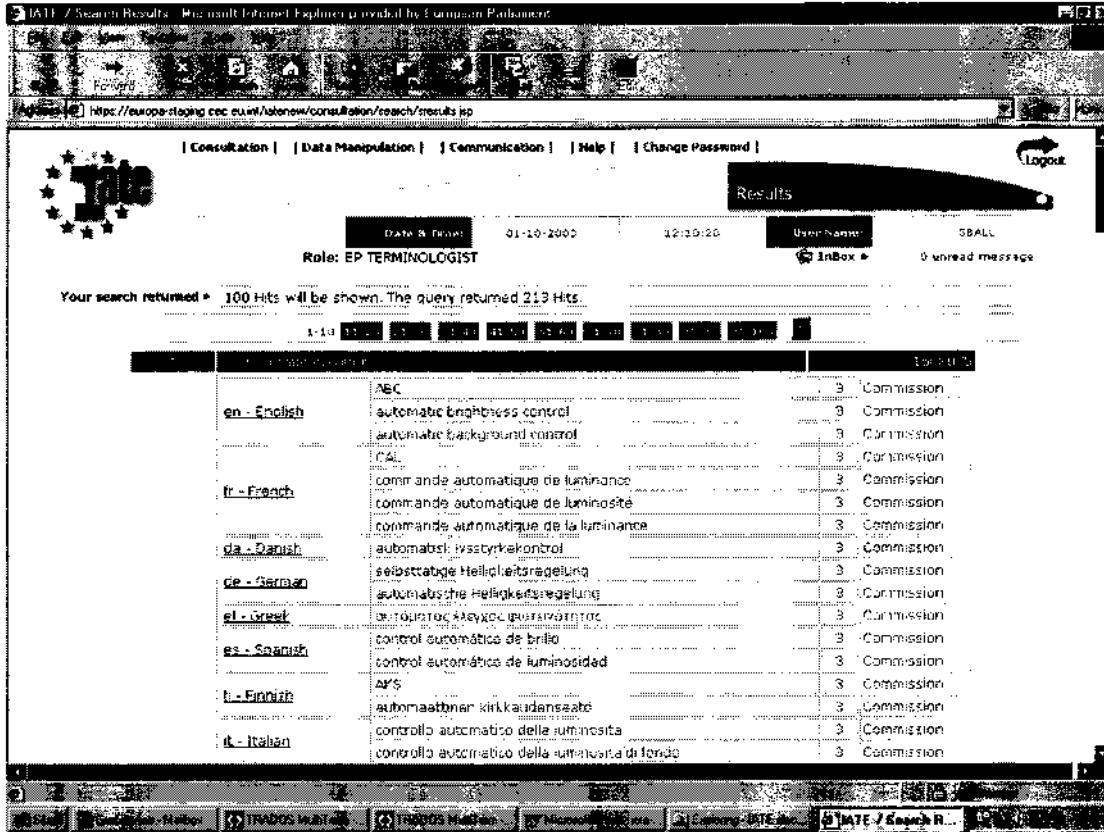


Fig. 3 The initial hitlist result

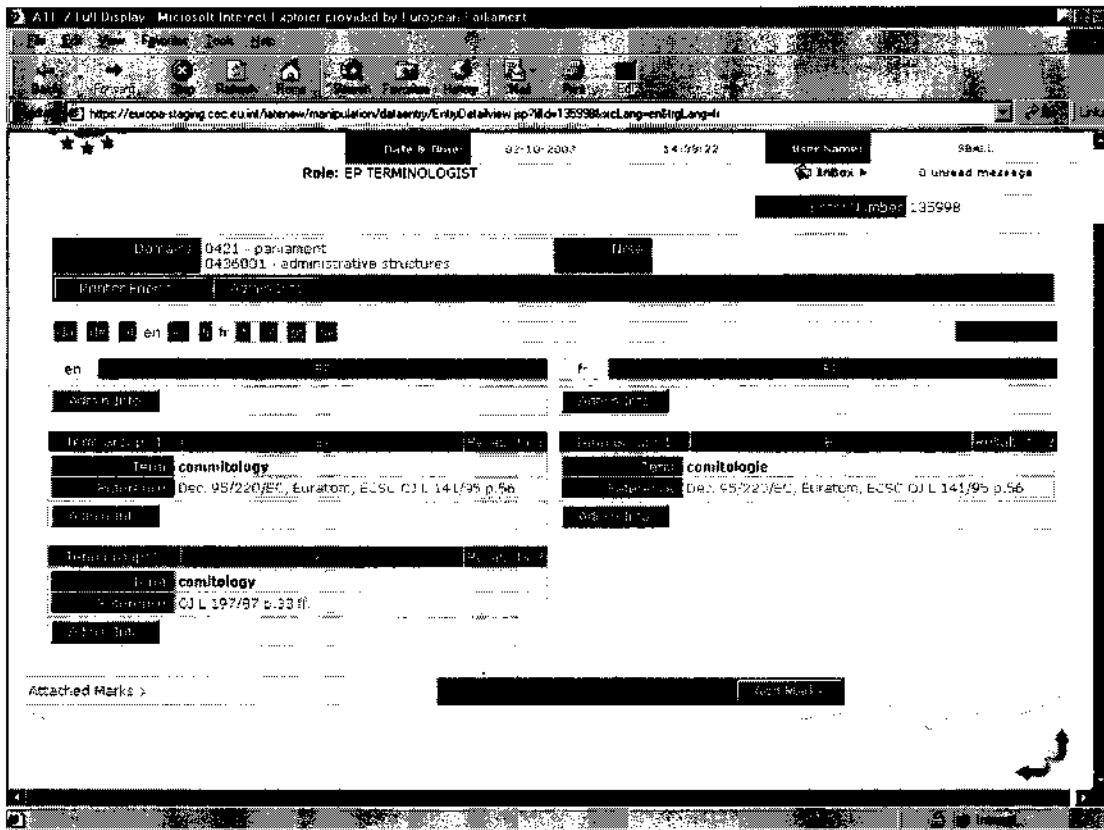


Fig. 4 Full display for "comitology"/"comitologie"

It is quite likely that for many searches the hitlist presentation alone will solve translators' problems, with the full display used for confirmation where they have doubts about definitions or partial solutions. However, terminologists will probably also want to use the system's more advanced search functions to manage records and improve their coherence<sup>11</sup>. This works in two ways. The first is a pseudo-SQL search via a rather intimidating blank interface, alleviated by a very useful help function and the possibility of checking the syntax of the query.

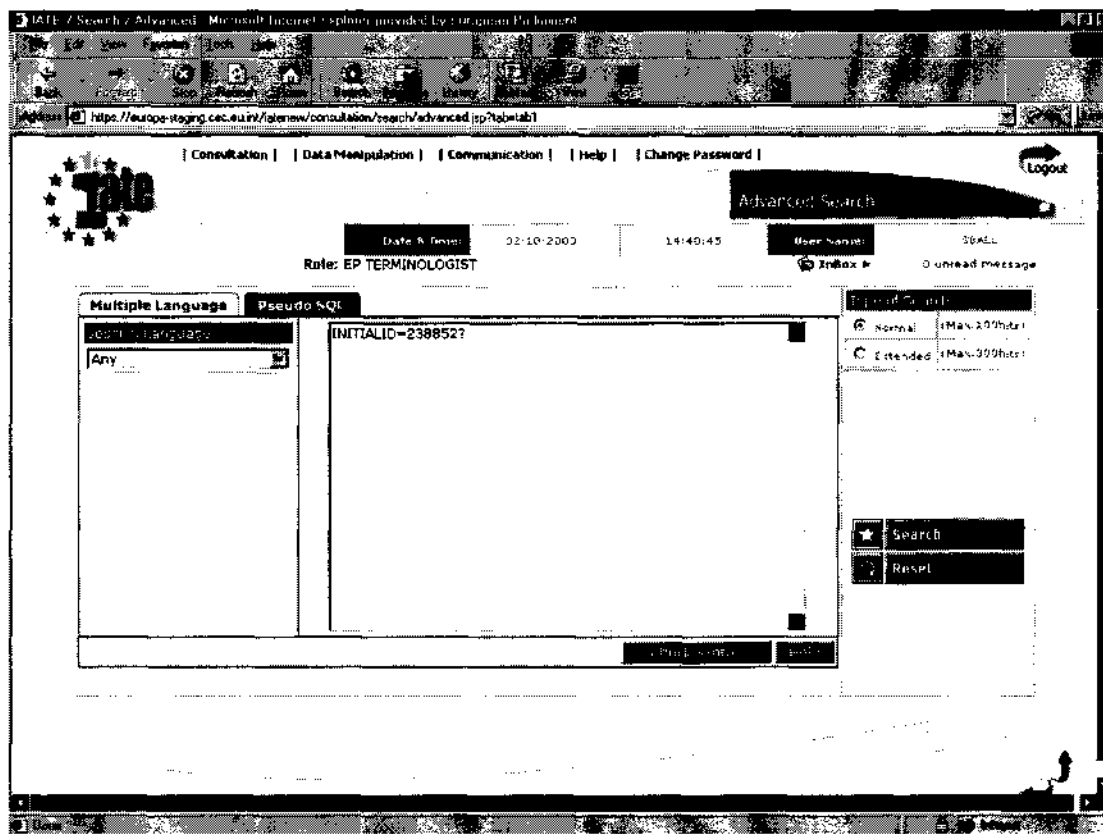


Fig. 5 Pseudo-SQL search for Euterpe entry 238852

Since the search argument for the query illustrated does not specify a language, the hitlist result is similar in colour to that for ABC, albeit rather different in precision since it returns only one hit. This is, however, only a very basic example. Search arguments can be combined with the Boolean operators shown in the next figure to give very refined results.

With the multiple-language advanced search screen shown in figure 6 it is possible to search for an entry with, e.g. “subsidiarity principle” in English and the equivalent in Danish (“nærhedsprincip”), but without “principe de subsidiarité” in French (x AND y NOT z).

<sup>11</sup> The advanced search is not available to guest users and thus not visible with the guest login.

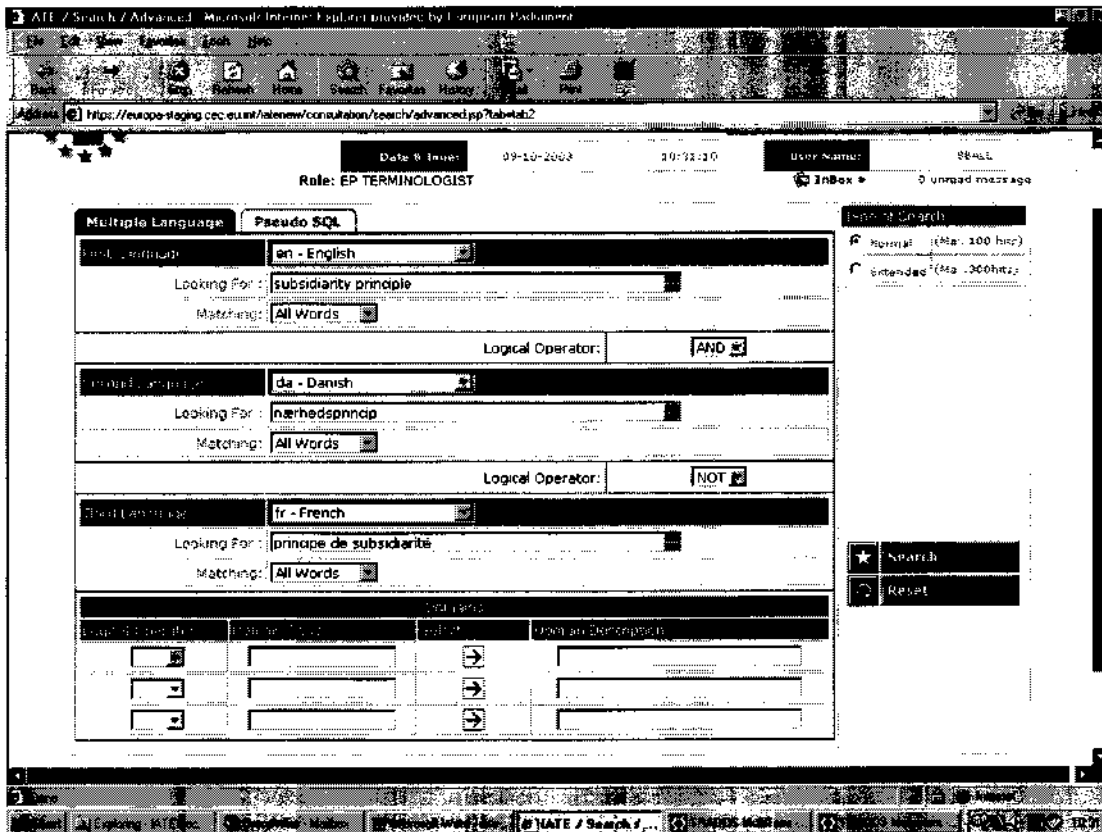


Fig. 6 The multiple language search interface

When translators or terminologists fail to find a solution on the database either because the concept is absent or because their target language is missing from the record(s) they retrieve (as would be the case for French in the example shown above), they are able to add a term or terms in their target language(s) to an existing record, or to create a new record in their source and target language, as relevant<sup>12</sup>.

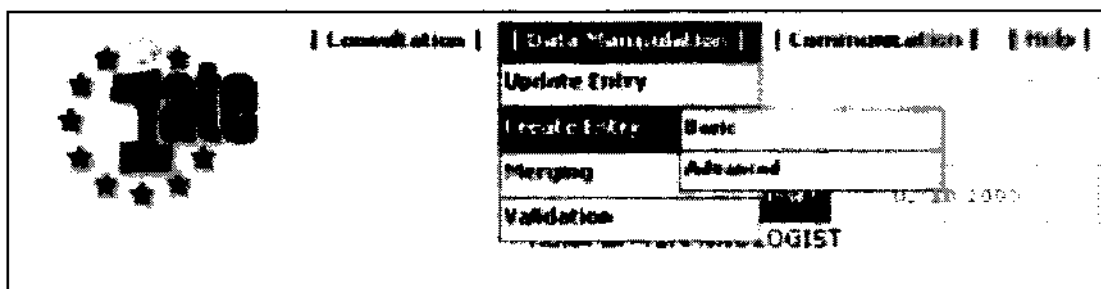


Fig. 7 Data-manipulation menu, showing options

With entirely new concepts, it is expected that translators, who after all are primarily concerned with getting on with their job, will mostly use the basic interface which requires only the selection of at least one subject domain and the entry of a source term plus context, or a source term plus abbreviation, or a source and a target term.

<sup>12</sup> Write access depends on the roles defined for users by the administrators of the participating institutions. At the EP all staff translators and trainees will be entitled at least to add a term or terms in their own language and to initiate a new record.

However, if users attempt to save a new record without a reference for the data they have entered, they will receive a warning message, which they can, of course ignore.

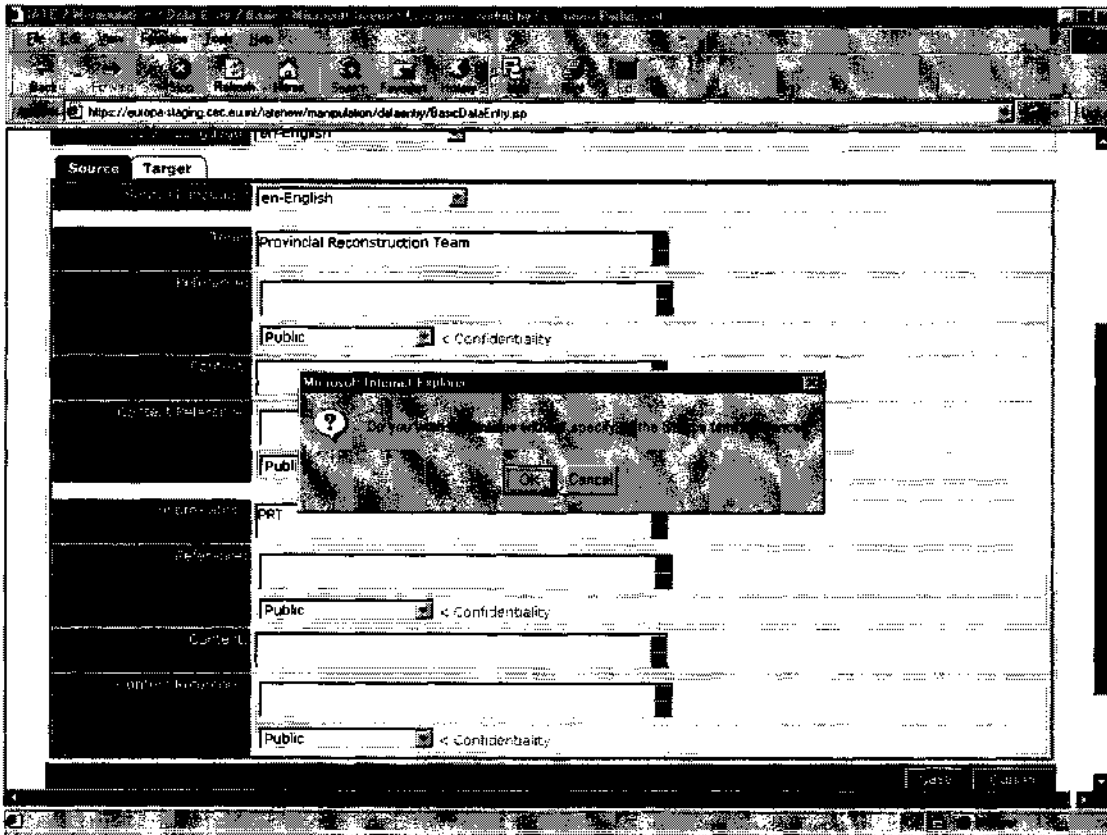


Fig. 8 Basic data-entry screen with warning message

The terminologists of the institution and language(s) concerned will automatically be notified that a new record has been created and it will be their job to liaise with users and make sure that records are as complete as is necessary to be generally useful. Where appropriate, they will use the messaging function described below to notify other divisions that they should complete the new record with their language(s). It is also expected that terminologists will in general work with the advanced data-input screen giving access to all the fields defined for the database<sup>13</sup>.

The IATE data structure is in fact on three levels<sup>14</sup>: concept, language and term, and the initiator of a new record with the advanced data-manipulation interface has to save each level as he goes. This makes it more suitable for experienced users such as terminologists (figures 9 and 10), although there is nothing to stop translators using it. However, there are certain tasks provided for in the system, primarily merging of legacy data or partial duplicates and validation, that are reserved for the terminologists' role.

<sup>13</sup> Translators will also work with the full interface to add terms to existing records, but this operation is less demanding than initiating a record for a new concept.

<sup>14</sup> Or four if the individual word level within multi-word terms is included.



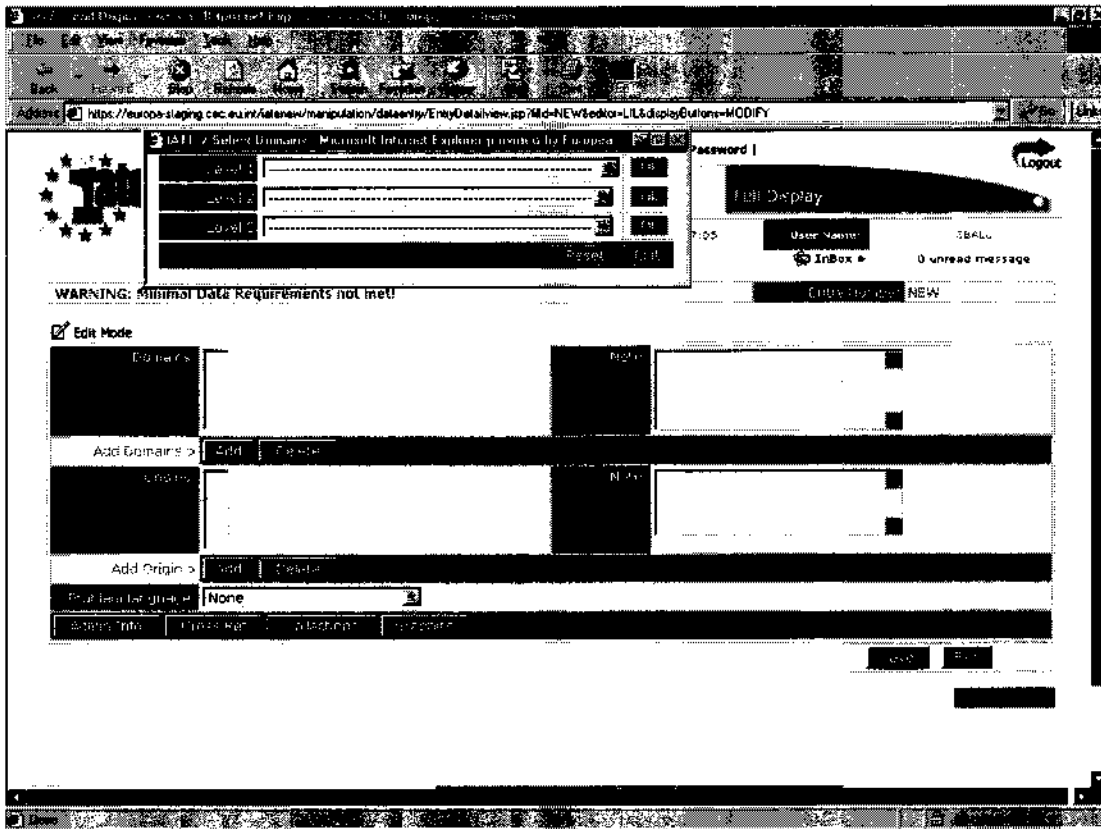


Fig. 9 First advanced data entry screen for a new concept

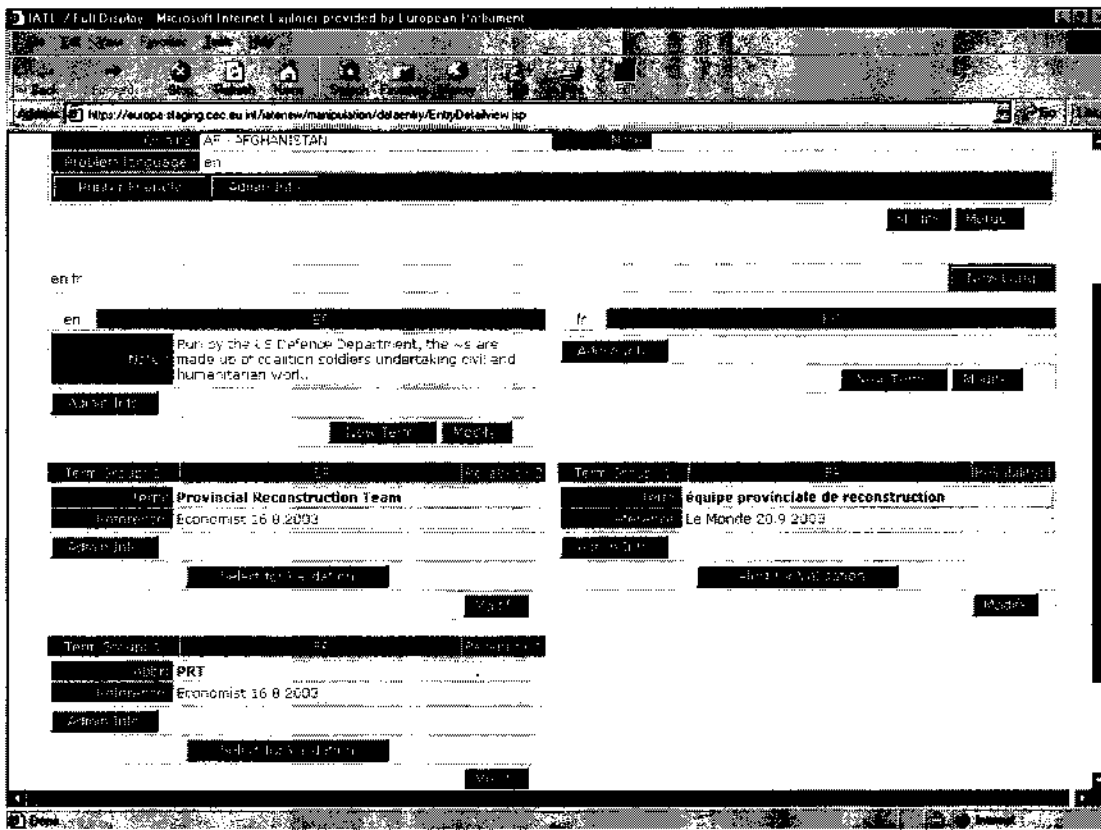


Fig. 10 A complete new entry in English and French for PRT

Merging entries is a typical example of a technical operation that will be confined to terminologists (or administrators, see below). It starts with the selection of two candidate entries for merging, the numbers of which are entered in the screen shown in figure 11. It then continues with an interface where the terminologist sees both records and has the option of selecting, deleting or concatenating data at the three levels (concept, language and term). Successful merging produces an entry like any other, so this feature will need to be demonstrated at T&C 25 if at all possible.

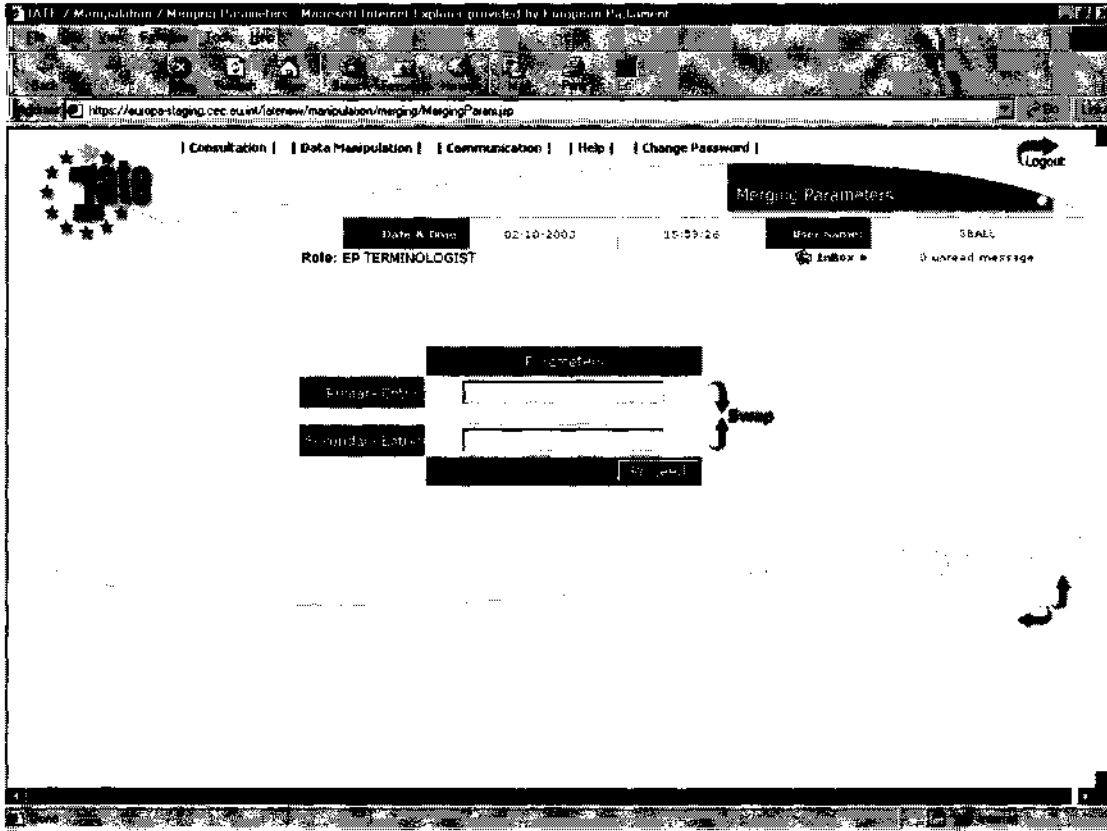


Fig. 11 The Merge operation opening screen

Validation is another area where terminologists based in translation divisions or central units will play a key role. For current EU languages it is a two-stage process, first for content (is this the correct term for the concept?), then for final verification that the entry complies fully with LATE standards before it can be seen by all users<sup>15</sup>. Terminologists can, of course, use their discretion to opt for final validation in a single stage, at which point the validation task is removed from the interface shown in figure 12. For non-EU languages the validation cycle is limited to a single stage (since at the time when the system was designed accession of new countries and thus the presence of translators into new languages seemed more remote than it does today), but as new countries accede to the EU, it is envisaged that their languages will move over to the full process.

<sup>15</sup> There are jointly agreed writing rules and a code of conduct for the IATE production phase; legacy data will be harmonised as much as possible.

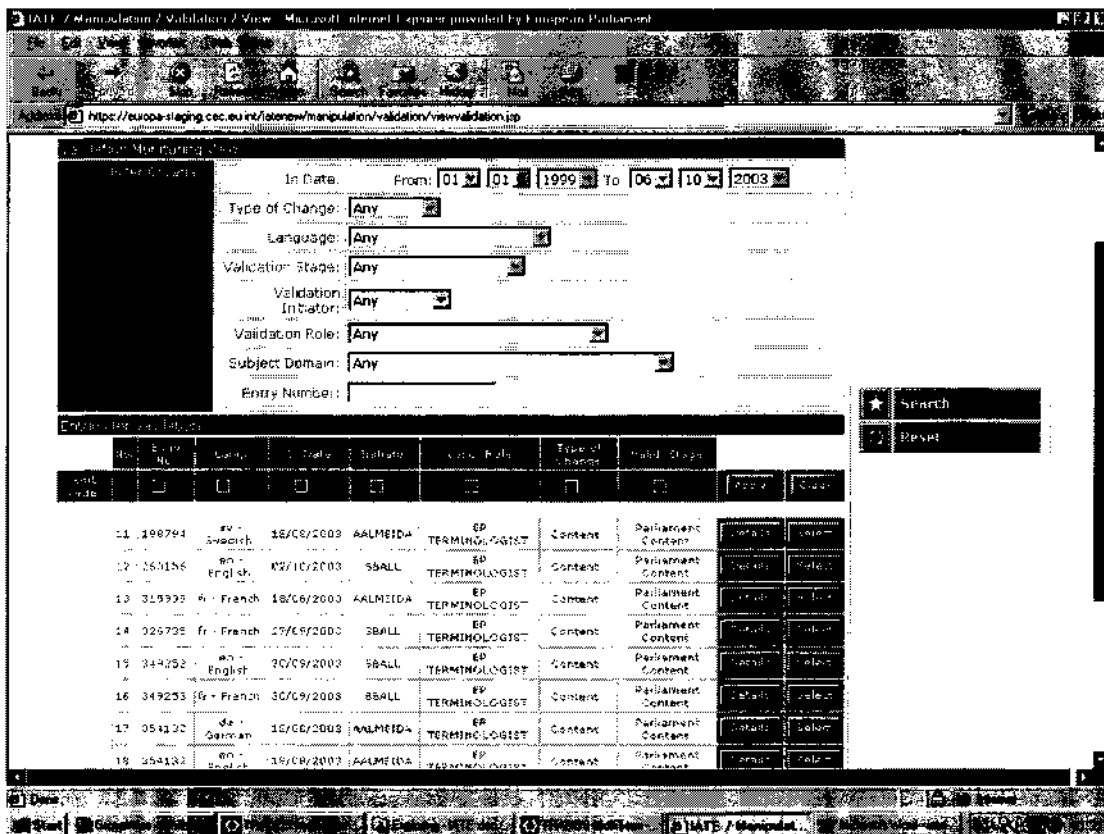


Fig. 12 The validation opening screen

While terminologists will be responsible for all data-related issues, it has been decided to appoint one or more administrators per institution, to take charge of technical tasks such as user administration, liaison with the system's central administration, exporting data for reuse in other applications (such as MultiTerm<sup>16</sup> or the Council of the EU's Glossary Production System) and, very importantly in the early days of the project, batch imports of legacy data, particularly for newly acceding languages. Administrators, of course, have full rights over all aspects of the system.

This all makes the IATE system seem very daunting - and in some ways it is - but there are numerous features built in to help users, in addition to the warnings shown in previous figures. One particularly helpful feature is the automatic detection of potentially duplicate records on the fly illustrated in figures 13 and 14 by the addition of a new English abbreviation (BPA) to an existing entry (bisphenol-A). The potential duplicates are shown in figure 13. They can be given a more enduring character as a "mark" attached to the record, as shown in figure 14<sup>17</sup>. (The situation shown in figure 14 is, however, purely for illustration. Users can - and should - choose to ignore the potential duplicates where they have checked that there is no real duplication.)

<sup>16</sup> There is a specific jointly agreed export definition for import into MultiTerm for use with the Trados Translator's Workbench.

<sup>17</sup> Marks are not, of course, visible to "guest" users. They can be thought of as evidence of "work in progress".

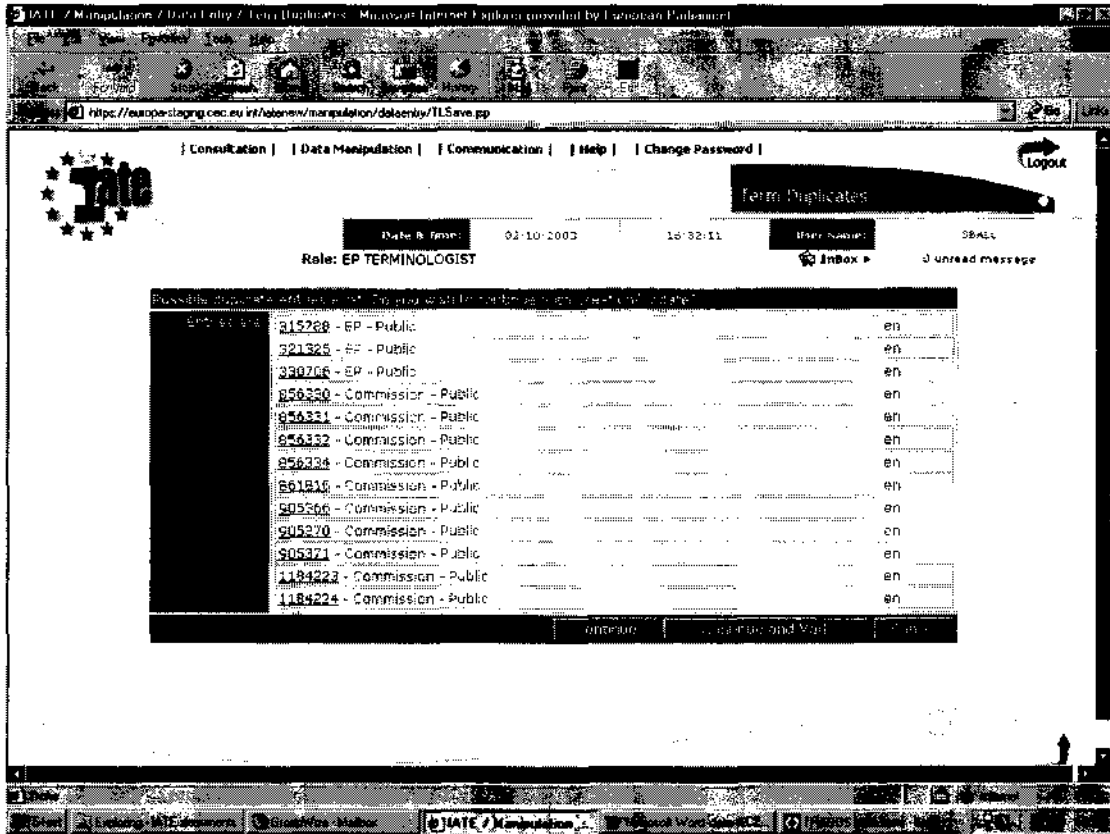


Fig. 13 Warning of potential duplicates after adding an abbreviation to a term

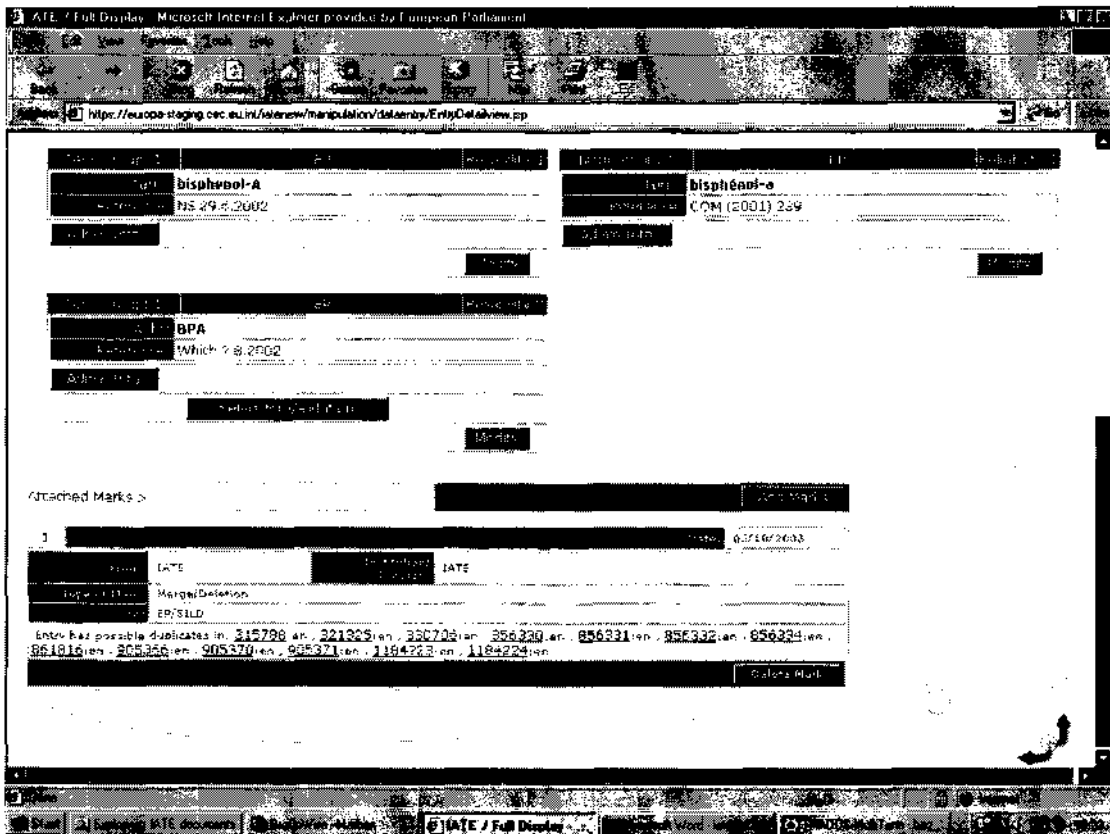


Fig. 14 The mark showing potential duplicates as hyperlinks

While marks are a form of internal communication about a particular entry, the messaging system is designed to facilitate communication between particular groups of users<sup>18</sup>. One instance where it would be useful is described above in the section on entry creation. It will also be a useful addition to terminologists' tools where they themselves create new entries, e.g. in our institution for new parliamentary bodies. Until now, when the EP's committees have changed we have been able to call up the text of the relevant decision and use copy and paste to create 11-language records in MultiTerm. Because of the layout of the text this was a quick and easy matter (except latterly for Greek). The new records were simply checked against the main database for potential duplicates using an internal string-matching programme and then imported.

However, in the more complex situation post-May 2004 where we will have 20 EU languages, this would be most unsafe and we will instead initiate a record in English and French, the "pivot" languages represented in the SILD central unit, and then use the messaging interface shown in figure 15 to inform EP terminologists of all other languages that they need to add the relevant term(s) in their language, preferably from the same official source<sup>19</sup>. It is also expected that users outside the central unit will use internal messaging to notify terminologists of entries which need to be merged (or split, although senior translators might be able to split an entry themselves, depending on the problem language and the other languages present).

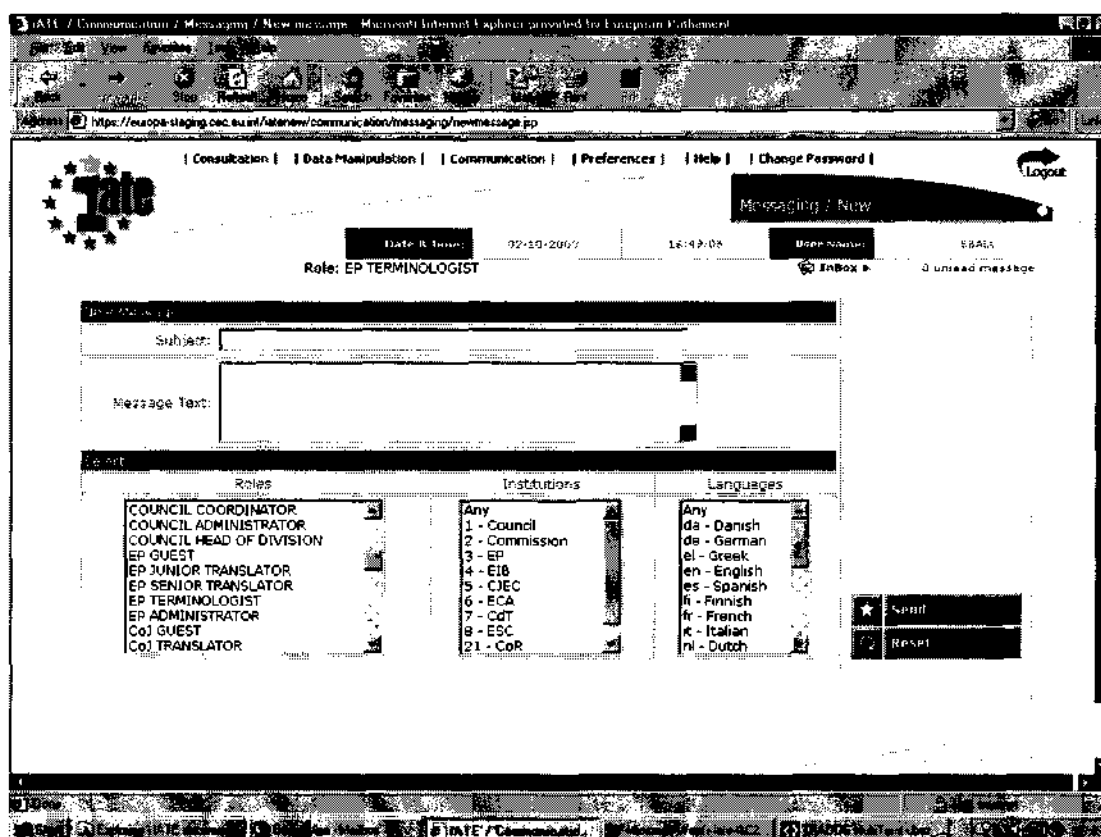


Fig. 15 The messaging system interface

<sup>18</sup> Messaging is only available to senior translators, terminologists and administrators.

<sup>19</sup> Multiple messages can be sent to similar groups, or a new group can be selected with Reset.

Technical features of IATE such as the marks and messaging interfaces may be restricted to experienced users, but there are some tools accessible to all levels of user, notably the possibility to create and save preferences as shown in figure 16. This is convenient because when the system was first demonstrated many of our translator colleagues criticised it because they felt it required too much use of the mouse. However, if you are a terminologist who regularly checks English abbreviations (as I do), loading the suggested “Abbrev-checking” preference takes only one mouse click, after which you can continue with the Enter key. If you are a translator who always translates texts in a set language from particular subject areas, the preference could even be loaded as part of your profile whenever you log on, so that you would not see irrelevant information. (unless too many searches returned “no hits” and you chose to widen your search).

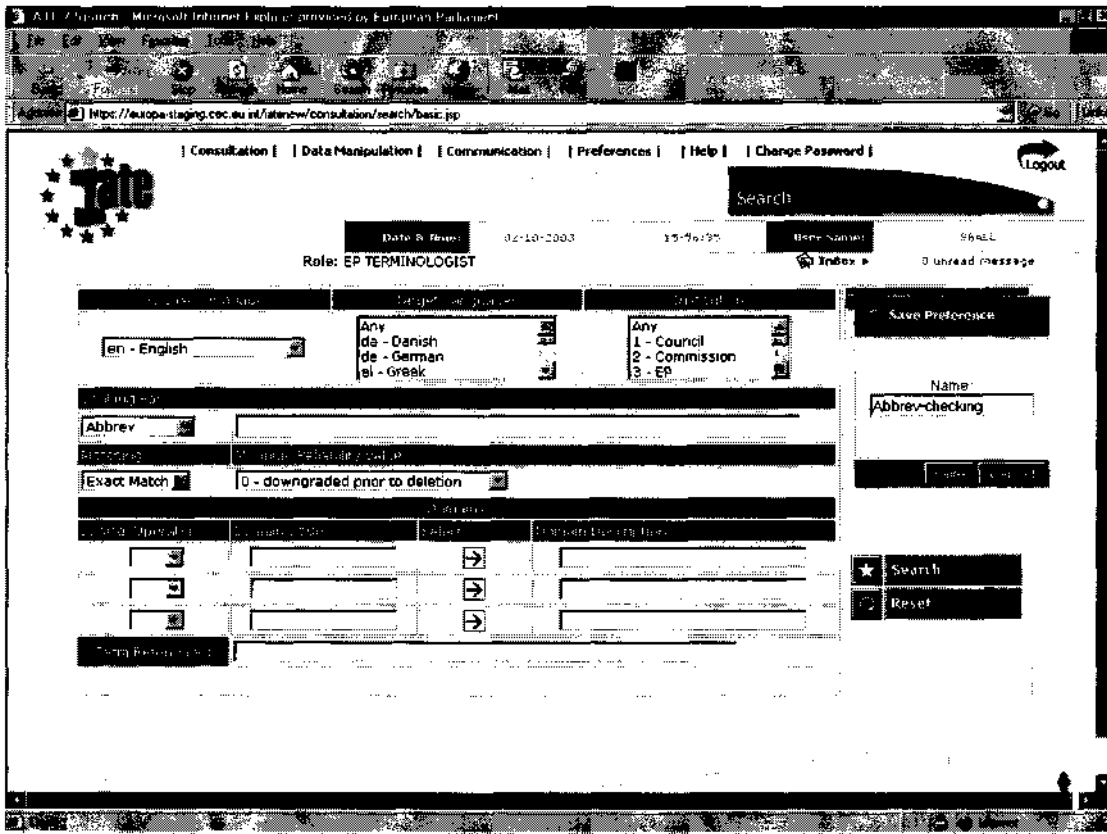


Fig. 16 Saving preferences

### ***Organisation and accompanying structures***

The IATE database is hosted by the European Commission’s Data Centre and managed by a team based at the Translation Centre. Astonishingly for such a complex undertaking, the central team comprises only 4.5 full-time equivalent posts (1 project manager, 2 web developers, 1 database administrator and ½ secretary) They are broadly responsible for leading and managing the project, on-going development, liaison with users and with the Data Centre, etc. One of their vital tasks at this stage of the project is acting as intermediaries for the participating institutions whenever a particular function has not yet been implemented in HTML and requires intervention in the Oracle tables, e.g. adding system users or changing user roles.

Clearly, with such a small central unit there is no intention of superseding the participating institutions' internal terminology teams with some grandiose supra-institutional structure. Instead there will be two interinstitutional groups, the Technical Coordination Committee and the Data Management Group<sup>20</sup>, under the overall control of the Coordinating Cell on Support for Translation and Interpretation which reports to the Interinstitutional Committee for Translation and Interpretation. This should not, however, be taken to mean that the institutions and agencies will continue to work in isolation as they did before. The tools and structures are in place for cooperation, harmonisation and a timely exchange of information. It would be nonsensical not to use them.

### ***Issues for the future***

One of the first issues to be tackled by the central unit has to be the reimplementation of any parts of the system not yet converted to HTML, so that user administration can be as flexible and decentralised as possible. More work will also have to be done on certain interfaces to improve their aesthetics and usability. However, technical issues like these are, for ordinary users, minor matters in relation to IATE's two main drawbacks, duplication of data and loss of the functionality previously offered by legacy systems, both of which will be far more problematical to handle.

Rules for the detection of perfect duplicates and their reduction were hammered out by the CST, Q&R's subcontractors, together with representatives of the participating institutions as part of the main IATE project. Unfortunately, because of the different classification systems used for data in the past and the differences in presentation of ancillary information such as references, etc., many duplicates have nonetheless been imported into IATE (remember "ABC"!), and the Data Management Group will have many difficult cases to resolve.

Equally, translators used to working with the Translator's Workbench and on-the-fly detection of terms present in a local MultiTerm database are less than totally impressed by having to connect to another system to look terms up there. To take account of such needs, administrators will make subsets of data available in MultiTerm (as suggested above) and, once we move to the production environment, users will have access to a link to IATE from Word, which should help to a certain extent. It remains to be seen whether further development will be needed in this area.

### ***Conclusion***

It would be wrong to be too self-satisfied at this stage. We have a long way to go inasmuch as we have to learn to work together over the long term and IATE still needs further improvement. As Dieter Rummel concluded in our paper for T&C in 2001, IATE like any other database system "remains only a tool"<sup>21</sup> but over the last three and a half years the project itself has shown the value of interinstitutional cooperation in the field of terminology. It makes sense, it improves consistency and it speeds up everybody's work. As cooperation between terminology services and other terminology actors in the future becomes a "given good", the database should

---

<sup>20</sup> The inconsistency of having one "committee" and one "group" is apparently coincidental.

<sup>21</sup>Rummel/Ball2001.

continue to improve. In short, joining-up our terminology resources shows that the terminology services of the EU are capable of joined-up thinking!

### ***Bibliography***

Almeida, A. (2001), 'Die terminologische Datenbank der Europäischen Union' in *Festschrift Joachim Göschel, Beiträge zu Linguistik und Phonetik*, Ed. Angelika Braun, Supplement to the *Zeitschrift für Dialektologie und Linguistik*, Stuttgart

European Commission, Translation Centre (1999), *IATE – Services for the Development of an Interactive Terminology Database System, Open Call for Tenders DGIII/99/050-IDA-101.02/01/IATE1*, Luxembourg/Brussels

Johnson, I., Palos-Caravina, M.-J. (2000), 'Validation and Quality Control Issues in a new Web-Based, Interactive Terminology Database for the Institutions and Agencies of the European Union' in *Translating and the computer 22*, Aslib, London

Johnson, I. and MacPhail, A. (2000), *IATE – Development of a Single Central Terminology Database for the Institutions and Agencies of the European Union*, Workshop on Terminology Resources and Computation, LREC 2000 Conference, Athens, Greece

MacPhail, A. (2000), *IATE - Inter-Agency Terminology Exchange*, Conference for a Terminology Infrastructure in Europe, Paris, France pp. 47-50

Quality and Reliability S.A. (2000), *IATE Project Validation Work Group Proposal*, Athens, Greece

Rummel, D., Ball, S. (2001), 'The IATE Project – Towards a Single Terminology Database for the European Union' in *Translating and the computer 23*, Aslib, London

Vidick J-L. and Defrise C. (1999), *Interinstitutional Terminology Database: Feasibility Study*, Atos, Brussels, Belgium, 147 pp.



## ***List of abbreviations***

CdT: Translation Centre for bodies of the European Union  
CITI/ICTI: Interinstitutional Committee for Translation and Interpretation  
CST: Centre for Speech Technology  
EP: European Parliament  
EU: European Union  
EuroVoc: European Vocabulary  
EUTERPE: European Parliament one-stop terminology management system  
IATE: Interactive Terminology for Europe  
OED Oxford English Dictionary  
OJ: Official Journal of the European Union  
Q&R Quality and Reliability  
SILD: Information-Technology, Language and Documentation Support  
Division  
SQL Structured Query Language  
TIS: Terminological Information System (Council of the EU)