

INFORMATION SOCIETY TECHNOLOGIES (IST) PROGRAMME



D14.2 – Final Report

Project Acronym: *ISLE*

Project Full Title: *International Standards for Language Engineering*

Project Ref. No.: *IST-1999-10647*

Project Start Date: *1st January 2000*

Project End Date: *31st December 2002*

Date of Preparation of Final Report: *12th February 2003*

Date of Revision of Final Report: *18th February 2003*

Editors

Nicoletta Calzolari
ILC-CNR & CPR
Pisa, I

John McNaught
Department of Computation
UMIST
Manchester, UK

Martha Palmer
CIS Department
University of Pennsylvania
Philadelphia, PA, USA

Antonio Zampolli
ILC-CNR, University of Pisa & CPR
Pisa, I

Table of Contents

1. Introduction	4
2. Background and Project Objectives.....	5
2.1. Background.....	5
2.2. Objectives	7
3. List of Project Partners	8
4. Organisation.....	8
5. Methodology.....	9
6. Focus of the Different WGs	11
7. Results	12
7.1. Results from the CLWG.....	12
7.2. Results from the EWG	15
7.3. Results from the NIMMWG.....	16
7.3.1. <i>Results from the US NIMMWG</i>	19
8. Conclusions	20
9. Deliverables List	22

1. Introduction

This is the Editors' introduction to the results of the ISLE project, which ran from January 2000 to December 2002.

As ISLE was a highly distributed project, involving numerous experts engaging in transatlantic cooperation to construct documents, exemplary resources and prototype tools of various types, yielding different results, the main aims of this introduction are to:

- enable the reader to understand the motivation for and organisation of ISLE;
- guide the reader by explaining the nature of ISLE results and the relationships among the various ISLE documents.

Taken together, the set of final document deliverables is referred to as the ISLE Guidelines. They represent the efforts of a wide scientific and industrial community to make recommendations for best practice in a number of areas of HLT and, where this has been possible, to propose standards in areas where broad consensus has been achieved. We emphasise that the standards described here are the first step in a long-term process: they represent the consensual view of the community on several issues and are to be seen as providing a basis on which further work can proceed.

The Editors thank all who have contributed to the ISLE initiative. There have been many kinds of contributions, all of which played an important part. All documents in the ISLE Guidelines have been planned, composed, discussed, revised and edited by teams of experts over three years, managed by the ISLE Working Groups and their Group Editors, and coordinated by the Coordination Team, thus it is invidious to mention individuals by name. Each document contains a list of those who contributed to it. Furthermore, and importantly, the wider community has been closely involved in commenting on preliminary drafts and on giving feedback at workshops and presentations at international events.

The Editors also wish to note that ISLE, being a joint EC/NSF project, presented both a challenge and an opportunity: a challenge at the organisational and managerial levels, given the numbers of people working in a highly distributed joint project and under two different funding, reporting and reviewing regimes, and an opportunity at the community level, as the project reached out to involve many experts, thus enhancing consensus-building. The project in fact surpassed expectations in terms of international cooperation, as it rapidly expanded to encompass contributions from and interaction with experts from Asian countries.

It is also highly noteworthy that many people gave freely of their time and effort in order to ensure the greater success of ISLE, both within and without the project, particularly our Asian colleagues.

The Editors furthermore note that industrial participation has expanded in ISLE, compared to prior EAGLES projects where it was already significant. The EAGLES-US International Workshop in January 1999 (in Pisa) was attended by numerous representatives of key HLT companies. All were supportive of EAGLES work and intended to become more closely involved. A workshop on EAGLES Evaluation aimed specifically at industrial and non-academic representatives held in Holland in April 1999 similarly revealed a strong interest in participating in future activities. This interest manifested itself in widespread industrial participation in ISLE from major US IT companies, namely Microsoft, IBM, Sun Microsystems Labs, AT&T Labs, Logos Corp, Systran Software, and GTE. In the EU, companies such as LexiQuest, Xerox Research Centre Europe, Lernout & Hauspie, Sail Labs, Compuleer, Sharp and ELDA were involved. Thus US and EU industry have had a substantial say in the development of LE standards within ISLE, which satisfies one of the main criteria for standards-oriented work, that industry should be a driving force.

In the following, we briefly characterise the ISLE project, placing it in context with respect to the EAGLES initiative both to emphasise the long-term nature of this type of work and to acknowledge the invaluable efforts that have taken place over the years within EAGLES which have helped give rise to, and determine the nature and form of, the ISLE project. We describe the motivation for the project and explain its methodology, its organisation, its objectives and its results, as embodied in the ISLE Guidelines and accompanying resources and tools. We also describe future work, given that ISLE is part of the ongoing long-term EAGLES initiative.

For information on managerial aspects, deliverables concerned with consortium issues, lists of project publications and events in which ISLE participants were involved, please refer to the ISLE project semestrial and annual reports, as appropriate.

2. Background and Project Objectives

2.1. Background

This project was funded as an Accompanying Measure for Key Action 3 “Multimedia Content and Tools” – Action Line 3.4 “Human Language Technologies”, under the Call for International Research Co-operation between the USA (NSF) and the EU (EC), within the Multilingual Information Access and Management initiative.

ISLE (International Standards for Language Engineering) is both the name of a project and the name of an entire set of co-ordinated activities that have galvanised the human language technology (HLT) field in recent years, under the aegis of the **EAGLES initiative (Expert Advisory Group for Language Engineering Standards)**, which has seen successful development and broad deployment of a number of recommendations and de facto standards via two prior projects, LRE EAGLES and LE EAGLES. This project thus reflected, in its objectives below, the objectives of a large community of researchers, producers and users of HLT. Although there were only a few named partners, these represented essentially the tip of an iceberg: the organisational model for ISLE, given below, was focussed around a few responsible partners who undertook to lead and co-ordinate work on the development of language engineering (LE) **standards**. This work was accomplished by a large variety of different parties, interacting in different ways in the HLT field, at a very low cost to funding bodies, yet produced results that are essential to the current and future development of the HLT field and that contributed directly or indirectly to most of the programmes and actions of FP5. ISLE emphasised work on an international scale, rooted however in ongoing EU and US RTD projects and National projects, towards standards of immediate, direct and critical relevance to the HLT industry and to users of the technology. This project was demand-led, as it responded to the stated needs of the HLT community (academia, industry, users), the EC and the NSF. It was also driven by the desire to foresee future needs and to plan LE standards actions appropriately to meet these as they arose. It is a simple fact that, without adequate provision for LE standards, EU and indeed national and international activities to enhance users' experience of and access to the information society will be severely curtailed. The following paragraphs place LE standards in their appropriate perspective, before then addressing the precise objectives of ISLE.

Many researchers, language engineers and technology planners have become aware of issues such as infrastructural resources, reusability, interchangeability, interoperability, and of their crucial role in facilitating the development of practical language technology products that respond to the needs of users. However, the implementation of a widely beneficial infrastructure in the human language technology (HLT) field relies heavily, as in other fields of technology, on the existence of common practices, guidelines, standards and compatible frameworks. Standards, whether these are de facto standards or national and international Standards, are the necessary key to true interoperability. With widely known and broadly accepted standards, interchangeability of language technology components becomes feasible; tools can be built to accept input or produce output in standard format; resources can be designed to a standard, or be mapped into a standard form; products of one type can be compared and evaluated, if they adhere to relevant standards.

The ongoing EAGLES initiative embodies attempts by a large community of language engineers to propose standards, guidelines and recommendations for good practice in several core areas of our field:

- very large-scale language resources (such as text corpora, computational lexicons and speech corpora, multimedia and multimodal resources);

- means of manipulating such knowledge, via computational linguistic formalisms, mark-up languages and various software tools;
- means of evaluating resources, tools and products.

Leading industrial and academic players in the Language Engineering field have actively participated in the definition of this initiative. Reports from EC Language Engineering strategy committees strongly endorsed standardisation efforts in HLT. There was and is a recognition that standardisation work is a necessary component of any strategic programme to create a coherent market, which demands sustained effort and investment, e.g.: ‘The EAGLES initiative to encourage the development of standards and protocols for all aspects of language and speech engineering should be continued as a priority.’ (Oakley, 1993, 73). In ISLE, the latest manifestation of EAGLES, we proposed to carry on working towards standardisation in our field, by developing a truly international approach to reaching broad consensus, especially in the crucial areas of multilingual HLT and of natural interaction and multimodality.

Progress in NLP and speech applications is hampered by a proliferation of different information formats, by variable linguistic specificity of existing information and by the high cost of development of resources. With the maturation of language technology, there come concerns about such matters as scaling-up, robustness, coverage, task suitability and accuracy.

EAGLES work towards de facto standards has already allowed the field to establish broad consensus on key issues for some areas — and as will be seen allowed similar consensus to be achieved for other important areas through the ISLE project — and thus provided a key opportunity for further consolidation and a basis for technological advance and expansion of knowledge. EAGLES previous results have already become de facto standards. To mention several key examples: the LE PAROLE/SIMPLE resources (morphological/syntactic/semantic lexicons and corpora for 12 EU languages) rely on EAGLES results and are now being enlarged at the national level through many National Projects; the ELRA Validation Manuals for Lexicons and Corpora are based on EAGLES guidelines; morphosyntactic tagging of corpora in a very large number of EU, international and national projects – and for more than 20 languages — is conformant with EAGLES recommendations. The work of EAGLES must be seen in a long-term perspective. This is especially true for any attempt aiming at standardisation in terms of International Standards. Moreover, successful Standards are those which respond to commonly perceived needs or aid in overcoming common problems. In terms of offering workable solutions, they must be based on some solid platform of accepted facts and acceptable practices. The HLT field is highly active and has met with success in several areas. This has indeed led to a welcome transfer of knowledge between different types of actors and languages, for example through EC sponsored programmes such as ESPRIT, LRE, MLAP and LE. There is a growing number of companies specialising in natural language processing and/or speech processing. However, although industry is capable of producing language engineering applications, the widespread development and adoption of such applications is threatened due to the lack of standards in the domain. EAGLES was set up to ameliorate this situation, through bringing together representatives of major collaborative European R&D projects in relevant areas, to determine which aspects of our field are open to short-term de facto standardisation and to encourage the development of such standards for the benefit of consumers and producers of language technology. This work is conducted with a view to providing the foundation for any future recommendations for International Standards that may be formulated under the aegis of ISO, and indeed one of the major outcomes of ISLE is that collaboration with ISO TC37/SC4 during the project has resulted in ISLE guidelines being taken as the basis from which to work towards such Standards.

The focus of the 5th FP, as concerns both HLT and the EU/US co-operation, is Multilingualism. It is therefore more than appropriate that in the area of Computational Lexicons the focus of the ISLE project was Multilingual Lexicons, while in the Evaluation area work concentrated on Machine Translation systems, long held to be the core multilingual application in the field, and one that has taken on enhanced importance in the DARPA TIDES initiative, with its emphasis on machine translation as a critical component in the chain of software tools allowing access to non-English language material on the Internet (stated objectives of the programme include translanguing capacity for 30 languages.). It is worth noting too that evaluation plays a central role in the DARPA programmes, and it is likely that the evaluation methods developed will be based on EAGLES guidelines.

However, new areas have arisen, dealing with natural interaction and multimodal/multimedia (NIMM) technologies. Over the last seven years, there has been a rapid increase of research activity in both NL and speech processing, in comparison to these new fields, whose technologies are still in their infancy. However, these new areas stand to benefit from early consideration of standardisation issues. For instance, in the gesture recognition community there was wide acknowledgement that a bottleneck exists, caused by the difficulty of sharing and exchanging gestural/multimodal data. There is little opportunity for re-use and there were no benchmark data sets for comparison of gesture and multimodal recognition algorithms. NIMM was thus chosen as an innovative area of concern for EAGLES, to lay the ground for early establishment of consensus. It is moreover important to note that these new fields are equally concerned with multilingualism and indeed with complex forms of communication where, say, several users of a system may each employ their own culture and language-bound methods of interaction, whether these be gestural, spoken, written, etc., and where there is a need to mediate between humans as well as the need to mediate between humans and systems that handle potentially massive amounts of information held in various forms and media and accessed in various modes. The broad spectrum of human communication abilities, including when mediated and expanded via computers, is essentially language and culture bound, thus it is of great moment to ensure that these new fields are encouraged to develop appropriately and for the applications developed in them to make a proper impact on society at large, by supporting them through early focus on standardisation issues.

As in the past with EAGLES, ISLE reached its objectives by building widespread consensus, thus avoiding—in the areas of concern—unnecessary duplication of work and dispersed effort. Even though EAGLES has full value by itself in a EU context, there was an even stronger need felt to drive towards widely acceptable standards through the framework of transatlantic co-operation, which is particularly important for topics such as multilingualism and the rapidly emerging area of natural interaction and multimodality. It is highly significant that there is already a well-established EU-US community in HLT where there is much contact and exchange of ideas, technologies and practices. Through EAGLES and the support of the EC, standards issues came to the fore as a major point of international concern. The work started in the EU spread outwards rapidly and became the focus of much discussion at international events. It was highly timely therefore to reach out, via ISLE, and bring in colleagues from the US to work with us in a common powerful drive towards standardisation in our field, which promises to have a telling effect on design, development, quality, reusability and interoperability of HLT products and resources to the overall benefit of all sectors of our multilingual, international society.

When considering the results of the project, the reader is advised to bear in mind that work on the various areas targeted by ISLE started from different levels of maturity with respect to recommendations and guidelines. Work on multilingual lexicons builds on results related to monolingual lexicons and corpora developed over many years, thus represents the most mature area. Work on evaluation builds on results obtained from prior EAGLES projects and on recent ISO standards. Work on natural interaction and multimodality builds partly on results from prior EAGLES projects (e.g. in relation to spoken language and to annotation of corpora) but has largely involved, given the newness of the topic, much groundwork to reach a level at which guidelines can be proposed.

2.2. Objectives

In summary, the objectives of ISLE were to:

- develop HLT standards from an international perspective;
- extend work of two previous successful projects on HLT standards, LRE/LE EAGLES, thus continuing necessarily long-term standards work, and tackling innovative areas where standards are strongly required;
- support HLT RTD and national projects, and HLT industry by developing, disseminating and promoting widely agreed and urgently demanded HLT standards and guidelines for infrastructural language resources, tools that exploit them and LE products;
- build on joint preparatory EU-US work towards setting up an international HLT standards oriented initiative;
- promote EAGLES, initially the focus of individual EU projects, as an internationally active body for HLT standardisation;

- contribute directly or indirectly to all IST thematic programmes.

As will be appreciated from the following, all objectives have been satisfied.

3. List of Project Partners

Participant Name	Participant Short Name	Country	Status ¹
Consorzio Pisa Ricerche	CPR	I	C
University of Southern Denmark	SDU	DK	P
Institut Dalle Molle pour les Etudes Sémantiques et Cognitives	ISSCO	CH	P
University of Pennsylvania – Computer and Information Sciences	UPenn – CIS	USA	*
University of Pennsylvania – Linguistic Data Consortium	UPenn – LDC	USA	*
New York University	NYU	USA	*
Information Sciences Institute – University of Southern California	ISI	USA	*

4. Organisation

On the basis of the previous EAGLES experience and of the proposed lines of action, the following organisation was adopted for ISLE:

- a co-ordinated, distributed body with an overall co-ordination mechanism (a Co-ordinator on each side of the Atlantic), secretariat and general editing/managing team;
- three Working Groups to work towards the formulation of standards and recommendations of best practice for the different areas/lines of activity:
 1. Computational Lexicons — **CLWG**,
 2. Natural Interaction and Multimodality — **NIMMWG**,
 3. Evaluation — **EWG**.

The WGs were co-chaired by a EU and a US member, and comprised members from both sides of the Atlantic. After ISLE got underway, Asian colleagues were invited to contribute to the work of the WGs.

Broad participation of experts in the WGs ensured that guidelines were formulated through a consensual bottom-up process in which the relevant scientific and technical aspects and positions were taken into account. Participation of industrial members in the WGs was essential to ensure industrial needs and

¹ C = Co-ordinator.

P = Principal contractor.

A = Assistant contractor.

* The following US research institutions:

- University of Pennsylvania – Computer and Information Sciences (UPenn – CIS);
- University of Pennsylvania – Linguistic Data Consortium (UPenn – LDC);
- New York University (NYU);
- Information Sciences Institute – University of Southern California (ISI);

contributed to the project under a NSF grant within the framework of the EU-US Science and Technology Co-operation Agreement (Oct. 1998), without being contractually part of the IST consortium.

Appropriate working agreements were concluded between the consortium and these US institutions.

Their technical contribution is described in the workplan (chapter 9 of the Annex I to the Contract for the project).

expectations were met, and as we saw above, industry was by no means slow to participate and ISLE acquired a strong industrial flavour. The EAGLES experience proved the utility of such a model and the existing community is now familiar with working in accordance with this model to good effect. EAGLES is like a Network of Networks (one for each WG), where major actors in the HLT community work together towards finding consensus on specification aspects and best practice in the different areas of concern.

The work of experts in the WGs for the carrying out of a specific task was implemented by means of sub-contracts from the relevant contractor (each WG was under the contractual responsibility of a partner/contractor). A substantial part of the project-allocated manpower required to accomplish the different tasks was normally therefore subcontracted. In addition, much work was provided through voluntary effort, offered by the participating organisations or by individual experts (in particular Asian colleagues, and senior researchers in participating organisations) for free, as in previous EAGLES.

Sub-groups (SGs) were created according to the different tasks, as the need arose. A WG acted as a steering committee for its SGs, which executed the day-to-day work in different areas.

The Technical Committee (TC), formed by the contractors, the WG chairpersons and Editors, the Central Editors and the Co-ordinators (all from both EU and US), monitored the overall technical and scientific coherence of the activities and results of the project. In particular, it formulated measures to ensure relevant synergies and interrelations among the various areas.

The Contractors, contractually bound to proper management and implementation of the project, shared the contractual responsibility for the overall operational, organisational and financial management of the project, and took the relevant decisions and measures. The contractor(s) responsible for a given workpackage ensured the appropriate timely performance of the work.

Users: companies —and representatives of National Projects and RTD projects —acted as validators in a similar scheme to that of the experts as members of the WG: participation as and when needed in the validation phase. A named individual acted as a contact point in a company; the company obtained early access to the specifications and data delivery, and tested these; it contributed to a feedback meeting or report.

For each WG, programmes of work were proposed such that each WG could progress at its own rate, in line with its objectives within its sphere of interest and the nature of its detailed work.

The Co-ordination WPs (WP13, WP14 and WP15) deal with the scientific and organisational management of the whole project, of the co-ordination aspects among the different WGs, taking care in particular of interrelations among them, of the TC and AB meetings, of general maintenance, evaluation, and dissemination aspects, of relations with other projects and external bodies.

5. Methodology

The quality of work in the approach to standardisation, implementation of the project and project results was continuously checked throughout the various steps of the methodology described below. The working process was based on regular communication and interaction with the research and development communities.

Work in the field of standardisation for HLT is in some respects different from normal project work, and parallel to it in others. In the course of the previous EAGLES work, the participants developed a general working methodology for the preparation of precompetitive recommendations for standards and guidelines which is applied in a flexible way according to the needs of specific topics.

This methodology of work proved fully adequate to reach the objectives of:

- pooling together results from different major projects;

- reaching consensus among the major actors in the different areas of concern of EAGLES;
- disseminating and making the community at large aware of the results;
- having the recommendations tested in external projects;
- getting feedback;
- having the recommendations/standards actually used in major projects.

This has been judged by the whole community the best and only way to set up the badly needed infrastructure of language resources in a coherent and harmonised way for all languages. It is a fact that the resources created in projects like LE-PAROLE and LE-SIMPLE — which have adopted EAGLES recommendations to design harmonised common models for lexicons and corpora for 12 European languages — are now enlarged through National Projects (thus implementing the subsidiarity principle) in at least seven EU countries. The EU is thus creating — with all these language resources based on a common model stemming from EAGLES — a very large standardised base of language resources throughout Europe. This would have been impossible without the role played by a standards-making project like EAGLES.

EAGLES aims at establishing de facto standards, emerging bottom up from consolidated practices. This has strong implications for the method of work, and on the structure of the ISLE groups, where major representatives of schools, applications, systems, etc., were represented. The already well-established EAGLES methodology, adopted here, has the following major steps:

- survey and inventory phase;
- discussion phase in WG meetings to reach and establish consensus;
- drafting of preliminary recommendations;
- validation actions for testing the practical applicability of the proposals (e.g. through preparation of small test resources);
- external evaluation: User Group, other projects, external experts, other languages;
- integration of feedback;
- definition of formal specifications and of operational guidelines;
- final recommendations;
- dissemination of recommendations and guidelines: Web, workshops, conferences, meetings with HLT projects, etc., to ensure and extend the consensus-based nature of EAGLES;
- maintenance of the guidelines (in this phase the US contributors will have also the chance to propose revisions to previous results, in order to make also previous standards truly international standards).

The different steps applied in a similar way to all the WGs; they can be clustered into phases normally ordered in time. The phases are the following:

- *state-of-the-art investigation*: identify relevant technical, scientific and engineering background, problems and existing solutions, products, application and system requirements, best practices;
- *elaboration of proposals for standards*: this is an iterative process which in many cases goes beyond the lifetime of a normal RTD project; it typically involves an increasingly broad public of specialists and professional users in the field, as can be feasibly done in EAGLES: this is also a means to achieve acceptance in the community at large;
- *validation of proposals*: this is in part a user-driven, in part a developer-driven action aiming at improving the proposals by applying them under controlled conditions in a real life situation; the results have an impact on the development of the proposals, on their dissemination and on practical applications; a small resource, a prototype tool, a parameterisation of an existing tool or another practically usable by-product is generated for the validation;
- *maintenance*: the standardisation process is a cyclic process of stepwise refinement involving interaction between a core group of proposers and an increasingly broader group of discussion partners, users and developers of products; a proposal for a de-facto standard is likely not to be “finished”, but at least “consolidated”, after a certain period of discussion; subsequently, e.g. applications to languages other than those initially considered may be added, and practical experience may lead to further additions and modifications. Even after full standardisation, regular maintenance is required to prevent obsolescence;
- *dissemination*: dissemination lies at the heart of standardisation activities, as much to gain feedback as to encourage uptake of the results. A multi-path strategy is called for, via numerous means, within

the available budget. Every advantage is to be taken of new technology and new media for dissemination. By its very nature, ISLE had in-depth collaboration with numerous related projects, initiatives and bodies at national and international level. Both the Co-ordinator and the WGs were involved in collaboration with outside bodies.

The different steps typically make different kinds of organisational set-ups necessary and involve in part different actors. Different types of work were thus carried out in different ways:

- survey & preparatory work,
- preparation of test data for validation of proposals,
- writing guidelines,
- which were all done both by partners/contractors and through subcontracts to other WG members;
- comparison of existing practices,
- discussion towards reaching consensus,
- decisions on actions and work to be done,
- which were carried out mainly through:
 - meetings/workshops of the WGs members/experts:
 - some EU-only or US-only,
 - some (at least one a year) of EU and US together,
 - at least one plenary Workshop of the three WGs together;
- workshops with users, other HLT projects, national projects and external experts, according to the Dissemination Plan.

6. Focus of the Different WGs

The three ISLE WGs targetted the three areas of multilingual computational lexicons (CLWG), natural interaction and multimodality (NIMMWG), and evaluation of HLT systems (EWG). These areas were chosen not only for their relevance to the current HLT call but also for their long-term significance.

For the CLWG, the objectives were to:

- extend EAGLES work on lexical semantics, necessary to establish inter-language links;
- design standards for multilingual lexicons;
- develop a prototype tool to implement lexicon guidelines and standards;
- create exemplary EAGLES-conformant sample lexicons and tag exemplary corpora for validation purposes;
- develop evaluation procedures for lexicons.

For the NIMMWG, working in a rapidly innovating domain urgently requiring early standardisation, the objectives were to develop guidelines for:

- the creation of NIMM data resources;
- interpretative annotation of NIMM data;
- meta-description of NIMM data resources;
- develop prototype tools for annotation of NIMM data resources.

For the EWG, the objectives were to:

- work towards definition of quality models for machine translation systems;
- maintain and update previous guidelines.

7. Results

7.1. Results from the CLWG

The CLWG has as its overall goal to establish a general and consensual standardised environment for the development and integration of multilingual resources. Earlier work in EAGLES had concentrated on monolingual aspects. Following on the success of this previous work, the multilingual aspect was targetted in the current project. The introduction to D2.2-D3.2 includes D1.1, introduction to the CLWG guidelines. The work of the CLWG evolved in a number of phases.

Firstly, a survey phase was initiated in order to gain knowledge of existing lexical resources, targetting specifically bilingual/multilingual, and semantic monolingual, lexicons. A variety of resources was covered, including industrial lexical resources, mainly via the participation of industrial partners in ISLE. The survey was later extended to a limited number of key Asian resources. This survey was initiated not only to learn about best practice in tackling the multilingual level of lexical description, but also to enable motivated selection of a number of critical areas either ripe for standardisation, or of present or upcoming importance to the field, where a push towards standardisation would foster an appropriate multilingual lexical environment to support expected or called-for key HLT applications.

Deliverables D2.1 and D3.1 (produced as a joint report) document the survey and the results of comparing the various lexical resources. The resources investigated included a number of major publishers' machine readable dictionaries, a number of computational lexicons (lexical databases) and a number of lexical resources that specifically support machine translation. Synoptic tables are given for each resource, showing how lexical information is distributed across each resource, and numerous examples, comparisons and discussions of how different resources treat specific cross-lingual linguistic phenomena are also provided, as are examples of how these resources handle sense distinctions. Major tendencies are brought out in conclusion, together with significant generalisations, and these form the basis for the development of guidelines. These deliverables were produced as a joint report as the decision was taken (and approved) to merge them, given that it was highly desirable that a single picture be obtained of how lexicographic practice dealt with bilingual/multilingual phenomena and relevant monolingual phenomena, and equally desirable to consider EAGLES-conformant lexical resources (e.g. PAROLE/SIMPLE lexicons) alongside other major resources.

Deliverables D2.2 and D3.2 (produced as a joint report) go into further detail on the role of the survey within the process of developing guidelines, but are principally concerned with the later phases of the CLWG's work, the work towards development of the guidelines themselves.

Following on from the survey, the CLWG launched two tasks to further prepare the ground for development of guidelines. These are documented in chapter 5 of D2.2-D3.2. These two tasks were aimed at identifying the lexical dimensions and the various types of information deemed most relevant for establishing multilingual correspondences, and thus the first task focussed on establishing basic notions for multilingual lexical encoding. These notions were catalogued and classified. Of these basic notions, many relating to the monolingual level of description had already been established in previous EAGLES work, insofar as the lower levels of linguistic description were concerned, with, as one might expect, coverage and depth of earlier recommendations being greater for the morphological and morphosyntactic levels than for the higher syntactic and semantic levels. This earlier work was now revisited to determine how it fitted within a multilingual perspective, and what may need to be modified or added. Happily, in the intervening years since the earlier EAGLES guidelines, significant lexical resources in several languages had been built up according to those guidelines, for example the PAROLE/SIMPLE lexicons, and ISLE was thus able to determine the implications for and effects of practical, large-scale lexicographic work based on those guidelines, for several languages. To further help in the establishing of basic notions, a number of sample lexical entries was constructed (see also discussion of D5.1 below), to investigate complex cases of multilingual correspondences. Monolingual entries were elaborated first, followed by multilingual ones. This strategy reflected the common approach of building monolingual resources independently, then linking them through multilingual transfer conditions. The second task focussed on sense distinctions, which are crucial

for successful multilingual lexical description, since cross-language links are essentially established between senses of different languages. Thus, an in-depth study of lexicographic sense indicators was carried out, i.e. the means by which lexicographers signal sense distinctions. This resulted in a database of sense indicators being built, with a tool to allow browsing and searching. This tool supports lexicographers wishing to establish translation equivalents, and was used during the construction of sample entries described above. Further studies were carried out on syntagmatic, collocational and contextual information used to establish cross-language correspondences or to indicate differences in meaning.

The bulk of D2.2-D3.2 is devoted to specification of the recommendations for the multilingual ISLE lexical entry (MILE). This is conceived as a highly modular and layered entry, such that MILE is able to respond to the differing styles and approaches to the lexicon adopted by various multilingual systems. As in previous EAGLES work, distinctions are made between obligatory, recommended and optional aspects of MILE (obligatory meaning: if you wish to be minimally MILE-conformant, you must respect this particular guideline). MILE is primarily seen as a meta-entry, i.e. it can act as a common representational layer for multilingual lexical resources to allow interchange of and ease of access to data, thus maximising reuse, integration and extension of lexicons. Instances of entries may and will differ in terms of the type and depth of information they include. An important consideration for MILE is that it can be used both for information-rich and less complex lexicons. Another crucial consideration is that MILE can be used to provide a means of communication and cooperation between those communities engaged in content-oriented description and access to services (semantic web, agent-based services, ontologies, content providers, ...) and those engaged in overcoming the language barrier, be it monolingual or multilingual (expert/non-expert communication, information extraction, knowledge management, machine translation, cross-language information retrieval, ...). Given these considerations, and others, CLWG has been working towards the definition of an object-oriented layer for lexical description, in order to foster more open and distributed lexicons. The lexical objects (MILE Shared Lexical Objects) associated with this layer can be used by lexicon and application developers at a high level of abstraction, thus aid in simplifying our recommendations and in improving their usability.

The model developed for MILE takes a transfer-based multilingual approach. However, direct and interlingual approaches can be accommodated within this model. An XML DTD (the MILE Entry Skeleton) has been written to formalise MILE.

The MILE Lexical Objects consist of three types of object: MILE Lexical Classes, MILE Lexical Data Categories and MILE Lexical Operations. The MILE Lexical Classes formalise the basic lexical notions discussed above, in a hierarchical organisation. The MILE Lexical Data Categories are instances of the Lexical Classes. They may be user-defined, or belong to a shared repository (a Lexical Data Category Registry). The MILE Lexical Operations are used to link syntactic and semantic descriptions and to provide constraints on such linking, and, importantly, to provide operations that can be used to carry out establishing of correspondences, tests and relevant actions at the multilingual level (including constraining and adding operations), to enable appropriate multilingual transfer to be specified.

A RDF schema for MILE has been developed, and example entries written using it, importantly demonstrating how a Lexical Data Category Registry can be exploited to provide off-the-shelf lexical objects which may be used as is, or serve as a basis for new or modified entries, thus enhancing reusability and ease of access.

A parallel phase, resulting in deliverable D4.1, involved work on the ISLE Lexicographic Station. This is a prototype development platform used to automatically generate a tool starting from the MILE DTD. As MILE was constantly evolving, the Station was designed to parse any Entity-Relationship model expressed as a DTD, in order to map it into a relational database and to provide a user-friendly interface covering the most common lexicographic requirements, including import from/export to external SGML/XML files. The main aims of the Station were to allow exemplification of MILE, use of external lexical resources, and the construction of exemplary entries according to MILE. A customisation module allows the user to overcome well-known shortcomings of DTDs. The web-based user interface covers the basic functionalities of query and browsing, import and export of data, encoding of new data, testing and validation of both data and the

lexical model, customisation, and various lexicographical operations involving type definition, class extraction and statistics. Discussion of the Station is also to be found in D2.2-D3.2.

Deliverable D5.1 concerns work to encode, in MILE, lexical data for a small number of entries and languages. This was carried out in two phases, for two purposes:

1. to provide a test-suite of possible multilingual transfer scenarios as input for development of the guidelines;
2. to enable stress-testing of MILE and to verify if it was sufficiently flexible to permit the representation of problematic aspects of mono- or multilingual entries identified during the course of the project.

Entry-writing in this manner revealed, as already suspected, that multiword expressions and collocations represented a critical issue. However, further experimentation with MILE lexical objects led to several proposals to handle multiword expressions, feeding into D2.2-D3.2, which furthermore demonstrate the flexibility of MILE.

Deliverable D5.2 concerns work to semantically tag a sample of text with sense tags. This task was carried out in collaboration with SENSEVAL-2, and involved the US ISLE Coordinator in setting up and running an evaluation Web site for SENSEVAL-2. An early decision was made to use EuroWordNet entries rather than the originally-intended SIMPLE (EAGLES-conformant) entries, as the former resource was more publicly available than the latter, and mapping from EuroWordNet entries to SIMPLE entries was deemed possible. By combining forces with SENSEVAL-2, ISLE was able to gather far more sense-tagged data, resulting from many different approaches, than was originally expected. Impressively, 93 word-sense disambiguation systems were submitted for evaluation, from 34 different research teams, which reveals the depth of interest in the topic. 12 languages were involved, and 3 different tasks were set (the distribution of languages over tasks varied, and not all systems were evaluated on all tasks, nor handled all languages). A workshop was held to discuss the exercise, although no detailed analysis of tagging results was possible due to time constraints. The detailed investigation of the relationship between lexicon and corpus, insofar as cross-linguistic realisations of the same word-sense is concerned, and the possibility of (semi-)automatic acquisition of multilingual correspondences from text corpora, remain open issues for future research. However, such research will have a solid basis of sense-tagged data on which to perform experiments, thanks to the joint ISLE/SENSEVAL-2 work, and will undoubtedly contribute in the near future to further enhancement of MILE.

Deliverables D6.1 and D6.2 report on work carried out to evaluate MILE-encoded data. D6.1 reports on an experiment to determine to what extent the multilingual aspects of two different machine translation systems, Rosetta (interlingual system) and Globalink (direct transfer system), the latter a representative of a broad class of commercial machine translation systems) can be expressed in MILE, carried out by an ISLE participant with strong industrial experience. This experiment was conducted using a preliminary version of MILE. The conclusion was that it was difficult to evaluate MILE as a proposed standard properly at that time, due to the then lack of a consolidated set of guidelines. (The current state of MILE and its documentation would now allow such evaluation). D6.1 however is an important document in that it contributed feedback to the MILE developers. It was noted that priority ordering of senses was a concern for machine translation dictionaries, as was a means of storing disambiguating phrases (used to prompt the user in interactive machine translation to aid in disambiguation when the system cannot resolve an ambiguity). Also, it was noted that there was a lack of clarity on how sameness of meaning was indicated, when using MILE in an interlingual translation approach. Globalink uses the Lexicon Interchange Format, and it was found possible, although difficult, to map LIF entries to MILE ones, principally because LIF makes no explicit distinction between source, target and transfer lexicons, and direct translation takes place. However, any mapping between LIF and MILE would only have to be implemented once. The more recent version of MILE is more amenable to participating in such mapping.

Deliverable D6.2, on user evaluation and feedback, reports on the main evaluation and feedback obtained from the international workshop organised by the ISLE CLWG in December 2002, at which MILE was presented, evaluation results were reported on and feedback obtained from ISLE-external sources, including Asian colleagues. This workshop also heard from representatives of projects in the areas of semantic web,

ontologies and industrial multilingual technologies. The former two represent important future constituencies for ISLE, given that, fundamentally, content is expressed, described, extracted and searched through reliance on computational lexicons. Critical reaction to MILE was positive and feedback provided in a spirit of aiding the further development of MILE to fully meet the needs of the wider community, especially regarding the needs of Asian languages. Further discussion of the feedback from Asian colleagues can be found in D2.2-D3.2.

Complementary discussion of D5.1, D5.2, D6.1 and D6.2 can be found in D2.2-D3.2, which also reports on associated work concerned with investigating the representation of noun compounds and support verbs (via XMELT) which fed into MILE, and on preliminary work towards handling multiword expressions in general. The main proposal for representation of multiword expressions rests on a notion of reuse of multiword expression patterns. It is noted that, for Dutch alone, this would lead to a reduction by a factor of 10 in the number of multiword entries to be stored in a reasonably-sized lexicon. Furthermore, D2.2-D3.2 includes an appendix on considerations for spoken language lexica in a multilingual context.

7.2. Results from the EWG

The current focus of work on evaluation is on methods and metrics for Machine Translation (MT) (earlier EAGLES work had looked at other application areas). However, this work is not being pursued in isolation, as MT is being used as a case study, to enable the later development of: a general theory about the methodology for evaluating HLT applications; and a general framework that can accommodate existing evaluation measures for specific HLT applications. The work carried out in ISLE therefore built on previous EAGLES evaluation work and, taken together with this previous work, is to be seen as part of a long-term effort to arrive at a general evaluation framework for HLT applications.

The EWG conducts its work in a different way to other EAGLES/ISLE working groups, in that it uses a series of practical workshops as its main means of refining the ISLE evaluation methodology and framework. These workshops are specifically organised to offer a forum for practitioners and industrialists to become trained in and to test out the current evaluation methodology and framework, by working on practical evaluation tasks, some chosen by the EWG, others proposed by participants. The results of these workshops provide feedback to enable evolution in the framework and the methodology. The EWG deliverables D13.1 and D13.2 report on the development of the Framework for MT Evaluation in the ISLE Project (FEMTI), D13.1 representing a draft version, D13.2 the pre-final draft version. Such is the interest that has been raised by the EWG work on machine translation evaluation that D13.2 has been already submitted as an article to a journal in the field, and thus appears in article form.

Early on, the EWG oriented its work around ISO/IEC Standards for software evaluation (ISO/IEC 9126 and ISO/IEC 14598). These Standards, however, remain at a level of generality, in that they provide a relatively abstract framework: the evaluator wishing to carry out an evaluation of a particular system in a particular context of use must further specialise these Standards for his own purposes. This is a complex task and in the absence of any framework or guidelines led to a plethora of different approaches to MT evaluation and little means of comparing across evaluations. In order to aid HLT evaluators, EAGLES set out to provide a means of guiding evaluators to develop their customised evaluations in a standard or at least highly principled way. Work on MT evaluation presented a challenge in that the MT domain is characterised by a severe degree of complexity and by many local contextual difficulties, which together have until now caused many different evaluation approaches and techniques to appear. ISLE recognised that there is no simple answer to the question of which is the best MT system. This implied that a successful approach would have to be one that could be parameterised. The questions then became ones of evaluation framework and design: how could a common parameterisable descriptive framework be developed, and how could the process of evaluation design for MT (and for HLT systems in general) be standardised? It is to be noted that ISLE does not propose new metrics, or attempt to automate the evaluation process, or to analyse the performance of human judges. Its concerns are higher-level, as, armed with a parameterisable framework and a set of principles to guide the process of evaluation design, an evaluator can become more productive, can produce evaluations that can be readily compared with others carried out within the same framework, and can more easily relate particular MT systems to contexts of use of interested clients.

Thus, ISLE took the ISO/IEC Standards as a starting point, and developed complementary extensions to these that enable practical MT evaluations to be carried out, via customisation of parameters in a principled manner.

In order to arrive at a motivated framework, the EWG initially investigated the various published evaluations of MT systems that have been carried out since 1979. This then led to consideration of the context of use of MT software. The ISO/IEC Standards relegate context of use to a lesser role, given their general nature, and give no overall indication of how to take this into account when evaluating a product. However, usability studies and the long experience of evaluation in the MT domain point to context of use being a strong influence on the quality model. Thus, a MT evaluator must be enabled to determine a particular quality model based on the expected context of use, in order to carry out an evaluation. This consideration led to FEMTI, which is based on three elements:

1. a classification of features defining a context of use in terms of users of the MT system under study, the task and the nature of the input;
2. a classification of quality characteristics of the MT system, with attributes/metrics appearing at the leaves; the upper levels correspond to the ISO/IEC quality characteristics;
3. a mapping from the first to the second classification, which defines or suggests the quality characteristics and associated attributes/metrics that are the most relevant for each context of use.

It is thus intended that MT evaluators can parameterise this framework according to the nature and context of use of the particular MT software they are evaluating. D13.2 gives only brief examples of the two classifications. However, the current state of the two FEMTI classifications may be viewed and browsed at <http://www.issco.unige.ch/projects/isle/taxonomy2/> where full information on the content of the taxons (whose titles only appear in D13.2) is given. The two classifications are encoded in XML, a single XML DTD being provided for taxons in both classifications at present. XSL is used to generate a readable version of the framework, enabling viewing of single taxons or of the entire framework (printer-friendly version). There is a facility to receive and log comments.

It should be noted that, as research is ongoing in the MT domain into the link between external quality requirements and internal features of a system, the EWG has not as yet been able to unify internal and external attributes under the six ISO/IEC top level quality characteristics, thus a separate classification branch has been provided for those internal attributes that do not lie under the six ISO/IEC quality characteristics.

In D13.2, an example is given of the steps involved in developing taxons, using the case of MT output quality (otherwise known as fluency). As demonstrated, this is a problematic area, and in general the classifications of FEMTI are not fixed in any concrete sense: they will continue to evolve as the field provides feedback.

As a result of EWG workshops, permission was obtained to scan, correct and disseminate an important, but hard to obtain, 1979 report on MT, the Van Slype Report. Also, work was completed on a small corpus of translations that served as input for workshop tasks.

7.3. Results from the NIMMWG

The NIMMWG is a young EAGLES group, compared to CLWG and EWG. Its main concern is to push towards early standardisation for a domain that is in rapid evolution, in which standardisation will serve to enhance development of large-scale reusable NIMM data resources, enable ease of description of such resources and promote development of portable and reusable annotation tools for NIMM data. In pursuing these goals, it builds on earlier EAGLES work on spoken language resources and on text corpora. Deliverable D7.1 provides an introduction to the work and results of NIMMWG.

The set of deliverables focuses on four aspects of NIMM: data resources, annotation schemes, tools to support annotation and metadata description of resources.

As with the other ISLE WGs, work was distributed between the USA and the EU. However, a slightly different approach was adopted to that of the other WGs, in that different sub-areas were targetted by the respective US and EU groups. This meant that there was little overlap and consequently limited transatlantic collaboration. Early on, a decision was approved that the US NIMM participants would not contribute to the European deliverables and vice versa. Nevertheless, the two sets of participants did interact in terms of joint participation in meetings and workshops. We here focus on the work of the EU participants of the NIMMWG, as the deliverables embody their results, however append short descriptions of the results of the US participants.

As the NIMMWG was targetting a domain that has only recently come into existence, and where little previous overview work had been carried out, it was firstly necessary to carry out substantial surveys. A first survey, reported in D8.1, covered NIMM data resources and included a strategic report (produced by ELRA) describing current and future user profiles, markets and user needs for NIMM resources. 64 resources world-wide were reviewed, a far larger number than previously surveyed. Resources were classified and commonalities and differences identified. It was found that re-use of these resources was rare and that many were poorly documented.

A second survey, reported in D9.1, covered annotation schemes for NIMM data resources and also surveyed best practice in annotation. 18 annotation schemes were investigated. It was concluded that there was much work still to be done towards development of an annotation scheme able to handle NIMM communication and data interchange in all their (cross-level and cross-modality) forms, at any level of relevant detail. There was, moreover, a lack of general-purpose annotation schemes within particular sub-areas of NIMM resources. Thus, NIMM ISLE work on annotation schemes was highly timely.

A third survey, reported in D11.1, covered tools, standards and user needs for annotation of NIMM data. 12 tools that supported annotation of some aspect of NIMM, including cross-modality, were reviewed. There was a marked scarcity of tools for NIMM data annotation, and most of the existing ones are research prototypes. Although existing tools strive to be general-purpose, it is currently difficult to attain such generality, mainly due to the lack of standards for NIMM data resources and for annotation schemes.

A fourth and final survey, reported in D10.1, covered best practice and trends in metadata description, looking at initiatives world-wide on metadata and headers for NIMM resources.

The main purpose of the above surveys was to yield input to allow the formulation of guidelines for the creation of NIMM data resources and their annotation (covering both annotation schemes and tools to support NIMM annotation).

Deliverable D8.2 proposes guidelines for the creation of NIMM data resources. Two types of guidelines are put forward:

1. guidelines regarding general resource specifications (legal aspects; modalities used; data organisation; and procedures, equipment, recruitment of subjects and assessment in respect of data recordings);
2. guidelines regarding data specification, for data in the three modalities of audio, image and video data.

The report also offers an introduction to validation of NIMM resources and validation criteria. In addition, it gives information on bodies and organisations involved in NIMM resource creation, dissemination and documentation.

The guidelines on general resource specifications are informal, and are meant to complement the work of the ISLE work on metadata, which resulted in a formal scheme for NIMM metadata and tools to support it.

The guidelines on data specification further emphasise the need to fully document the contents, procedures, tools, standards and particular modalities of the resource being created.

Deliverable D9.2 summarises the results of D9.1 and then proposes guidelines for multilayer annotation of uni- and multimodal data. It covers the following aspects:

- how to create NIMM coding schemes;
- how to document coding schemes;
- how to represent coding schemes and annotations in a computer readable format;
- how to locate, select and evaluate an appropriate coding scheme;
- how to adapt an appropriate existing coding scheme.

D11.2 and D11.3 provide respectively requirements for and implementation of a prototype annotation tool for NIMM data. The main requirements identified were:

- a flexible and open architecture which allows for easy addition of new tool components (a modular workbench);
- separation of user interface from application logic and internal data representation. The internal data representation should be separated from the user interface via an intermediate logical layer so that the former two layers can be modified separately;
- transcription support and annotation support at different levels of abstraction, for different modalities, and for annotating relationships across levels and modalities;
- to the extent that it is possible to (semi-) automate NIMM annotation processes, this should be supported by the toolset. Similarly, to the extent that it is possible to (semi-) automate NIMM data analysis, this should be supported by the toolset. Automation should be supported in two ways: (i) via the possibility to add (through an API) additional components for automatic annotation and data analysis, and (ii) via the use, as far as possible, of standard(ised) data formats, allowing easy importing and exporting of (automatically created) annotations;
- powerful functionality for query, retrieval and extraction of data from annotated corpora; tools for data analysis, possibly including statistical tools;
- adequate support for viewing and listening to raw data;
- adequate visual presentation of annotated data;
- easy-to-use interface. In general, the tool interface should support the user as much as possible, be intuitive, and as far as possible be based on interface standards which the user can be expected to be familiar with;
- support for easy addition and use of new coding schemes and for defining new visualisations of annotated data (e.g., presenting annotations based on new coding schemes);
- possibility of importing and thus reusing existing data resources via conversion tools;
- possibility of exporting by means of conversion tools, coded data resources for further processing by external tools;
- most importantly, perhaps, the tool must be robust, stable and work in real time even with relatively large data resources and complex coding tasks.

These requirements are embodied in the implementation represented by D11.3. This work was based around the MATE workbench, and also formed the starting point for work in the NITE project. It is to be noted that similar, complementary work was carried out by the US NIMMWG, which gave rise to the AGTK suite (see below).

Deliverable D10.2 reports on the final version of the NIMWG guidelines for metadata description. The ISLE metadata initiative (IMDI) has developed a complete metadata infrastructure for the description, discovery and management of multimedia language resources. Central to IMDI's concerns was the creation of a metadata vocabulary appropriate for HLT requirements. An entire sub-proposal concerning metadata for computational lexicons was elaborated, as a result of interaction between IMDI and the ISLE CLWG. The format of metadata descriptions is XML based. The framework developed allows flexible creation of hierarchically-organised metadata descriptions, which may moreover map to distributed, remote language resources. Deliverable D10.3 reports particularly on the development of an Internet site with portal that supports editing, browsing and searching of metadata descriptions. Some 15,000 metadata descriptions have been made available. Interoperability is ensured between IMDI and the standard Dublin Core metadata set, such that metadata service providers using the Open Archives Initiative protocol can harvest records from the IMDI universe (although some information loss may be expected). IMDI advanced its work through a

number of workshops at key international events. An IMDI showcase covering metadata from 6 European organisations was organised for the opening ceremony of the European Year of Language, in 2001. IMDI has been particularly successful in that its results have been quickly adopted and have fed into both the ECHO project and the recently started INTERA project that will establish a European language resource area based on the IMDI metadata set.

7.3.1. Results from the US NIMMWG

A brief description of US NIMMWG results is now provided, given that, as mentioned above, the US NIMMWG members did not participate in the EU deliverables. US NIMMWG work concentrated on the 3 subareas of:

1. Spoken Language;
2. Gesture;
3. Discourse.

The spoken language work focussed on OLAC, the Open Language Archives Community, which has the goal of laying the foundation of an open, web-based infrastructure for collecting, storing and disseminating the primary materials which document and describe human languages. This initiative has seen strong contributions in terms of leadership and organisation by US NIMMWG members. The OLAC project has continued to generate best practice guidelines for resources. Its activities are summarised in postings to OLAC General: <http://lists.linguistlist.org/archives/olac-general.html>.

A major workshop was held, organised by US NIMMWG members, to revise three proposed standards — the OLAC Metadata Set, the OLAC Process document and the OLAC Protocol. Other work was pursued also on finalising controlled vocabularies, giving feedback to each participating OLAC archive on its use of metadata, and reviewing the services on the OLAC and LINGUIST sites. The outcome of the workshop includes the above three proposed standards, which are currently being reviewed by the OLAC community, several best practice recommendations, and a variety of informational and implementational notes. All of these address the need to document spoken language resources, along with language resources more broadly.

Another major contribution of the US NIMMWG members working on spoken language was in the area of annotation tools: a suite of software components, the Annotation Graph Toolkit (AGTK), was developed, for building tools for annotating linguistic signals, including time-series data, which document any kind of linguistic behaviour and do so in a layered way, for multimodal annotation. AGTK is available from <http://agtk.sourceforge.net>.

The gesture work saw US NIMMWG members participating in an IEEE international workshop on cues in communication, which advanced discussion on annotation standards for gesture. The main thrust of the group's work on gesture involved studying annotators to determine how they spend their time in annotating (mainly on encoding human-readable descriptions of arm movement), leading to development of a new annotation scheme, FORM. This is the first of its kind in that it encodes gesture information kinematically and can be used on pre-existing video data (thus avoiding expensive laboratory work). 10 minutes of exemplary data have been annotated with FORM. Preliminary inter-annotator agreement results are promising. These data have also been augmented with Prep-Stroke-Annotation data. Further annotation has been carried out to provide a corpus of FORM motion-capture data, with a view to allowing further verification of the scheme. Also, about 1 hour of FORM video data has been produced, covering a variety of types of gestures, including natural chimpanzee gestures. All annotation has been hand checked for accuracy. A guide to best practice using FORM has been produced, together with a description of the annotation and a training video for annotators.

The flexibility of FORM has also been demonstrated, as experiments have been conducted on mapping FORM data to a parameterised animation system.

As for the discourse work, the CMU corpus of human-human dialogues between a travel agent and a customer was selected for discourse/dialogue annotation, using dependency tree annotation. Further dialogue

act tagging was carried out on the HC Communicator corpus of approximately 2000 dialogues. A 40 dialogue subset of this corpus was labelled for concepts, misunderstandings, initiative and dialogue act tags. A small corpus collection in the restaurant information domain using a Wizard of Oz paradigm has been completed. The audio and transcriptions for that corpus remain to be digitized.

An ISLE Workshop on Dialogue Tagging for Multi-modal Human Computer Interaction was held as a joint venture between US and EU NIMMWG groups. While previous workshops on dialogue tagging have focused on developing a standard tagset that could be useful across a broad range of applications and domains (e.g. DAMSL), this workshop explored the hypothesis that progress in the field will be achieved more rapidly by identifying specific reference tasks to which a tag set is relevant. The participants in the workshop proposed and explored tagsets that have clear relations to important multimodal human computer interaction tasks, and where tagging can be shown to support algorithms that demonstrably improve performance on such tasks. The aim is to move away from generic proposals of tag sets, and instead to propose tags that are relevant to particular tasks. The workshop proceedings are available at <http://www.research.att.com/~walker/isle-dtag-wrk>. The conclusion of the workshop was that there are many application and domain specific features of coding schemes which have to be developed with each new application domain. However, there was broad agreement on the utility of having a general reference coding scheme such as DAMSL that other coding schemes could be related to.

8. Conclusions

As may be appreciated from study of the aforesaid results, ISLE has met its overall objectives. The project has moreover engaged with many national and international projects and initiatives (see Annual Reports and Semestrial Reports for details) and has been represented at many international events. The rate of publication of the NIMMWG is particularly impressive in terms of raising awareness on both sides of the Atlantic, and internationally. The CLWG has been active in establishing and exploiting links with Asian countries in order to further refine MILE. The EWG has developed strong international interest in its work, through its workshops. For all groups, there has been a highly satisfactory level of industrial participation and interest. As may be expected from a project of this kind, progress towards concrete recommendations has been variable. The most concrete recommendations are to be found in the results of the CLWG and EWG, and of the NIMMWG IMDI subgroup. Other NIMMWG subgroups have produced essentially first drafts of recommendations, for creation and annotation of NIMM resources, due to the different nature of their domain which has not benefited from prior work on recommendations for the most part, apart from the earlier work of EAGLES in the spoken language domain and in text corpora. CLWG and EWG were already building on solid foundations and the nature of their recommendations reflects this.

It is also however true that one cannot consider even the more mature recommendations to be final. Given the short time period of the project, the limited funded person-power and the fact that ISLE is seen as another step along the route already mapped out as a long-term voyage for the EAGLES enterprise, we fully expect our current recommendations to undergo refinement and modification as the wider community reacts to them. We also expect EAGLES/ISLE results to be taken up as appropriate by ISO TC37/SC4, recently founded with the aim of taking recommendations from the HLT community and turning them into ISO Recommendations and eventually Standards. The coming into being of this committee has in itself been proof of the success of the EAGLES/ISLE initiative, in that ISO now judges it timely to proceed towards standardisation in the HLT resources field and related areas, a judgement made on the strength of support in the community for EAGLES/ISLE guidelines.

In general, EAGLES/ISLE is also already looking to the future, as the following explains.

The CLWG is exploring the notion of new-generation, open, distributed and shared lexical resources. It recognises that a close relationship exists between computational lexicons, content processing, ontologies and the Semantic Web, in that all are concerned not just with format and encoding, but also with content, and

content expressed in many languages. This close relationship is however implicit rather than explicit: the CLWG has taken a first step, at its end of project workshop, to involve the other domains mentioned above in discussions in order to emphasise the importance and central role of computational lexicons. MILE is viewed as the beginning of a common platform on which reusable, open, distributed and shared lexical resources may be built that will allow interoperability of content. However, major efforts are implied here, and thus a suitable infrastructure is required. An expression of interest entitled ‘European Lexical Infrastructure and Technology’ (ELITE) has been submitted to the EC within the 6th FP with a view to pursuing this vision in the future. In the meantime, EAGLES will continue its work, coordinating efforts to, for example, further evaluate MILE to determine how easy it is to map different lexicon models to MILE. Mappings to and from OLIF resources, (Euro)WordNet, PAROLE/SIMPLE, ComLex, FrameNet lexicons and others will be examined in this respect, as participants in the end of project workshop undertook to become involved in helping to test MILE. Initial future work will be focussed on defining a methodology to allow such widespread testing to be undertaken, which will also help to raise the credibility of the ELITE initiative. Further future work is planned to deepen analysis of sense-tagging results from the ISLE/SENSEVAL-2 initiative.

Importantly, also, future work will consolidate and extend cooperation with Asian countries. ISLE guidelines will be extended to deal with the many particularities of the universe represented by Asian languages. It is highly noteworthy that the ELITE initiative was strongly supported by many non-EU entities, such as the newly formed Asian Federation of Natural Language Processing Associations (AFNLPA), the Academia Sinica and KORTERM (Korea). Within the EU, the newly formed Coordination Group of National Projects on Language Resources in Europe (ENABLER, currently representing 13 European languages) has declared a deep interest in assessing and implementing EAGLES results. Thus, experiments with mapping between existing lexicons and MILE, and growing cooperation with key bodies and initiatives on a world-wide basis, will pave the way towards the creation of an open, distributed lexical infrastructure (ODLI), which will represent a new paradigm of distributed lexicon creation and maintenance, that will strongly support the full realisation of the Semantic Web. The work of EAGLES CLWG towards elaboration of consensual standards and best practice thus continues in the foreseeable future with a definite direction and an ever-widening circle of interested and motivated participants.

In future work on MT evaluation, the EWG expects to tackle classification of the purposes of evaluation and of the objects of evaluation. They also expect to extend the work that led to FEMTI to other fields besides MT, taking into account that this is more obviously feasible where specific needs for specific contexts of use can be identified for some system. The current work on MT is considered but a start by EWG, a first step on the way to greater understanding of the complex nature of MT evaluation. Further work is required on the strengths, weaknesses and correlations of various metrics for MT evaluation, which demands a dedicated programme of systematic comparison. Thus, for each taxon in the quality characteristics classification, it is ultimately necessary to know, given the purpose of some evaluation, the most relevant attributes, the most appropriate associated metrics, the work and cost involved in applying each metric, and the score level that should be considered acceptable.

Future work on the XML-based FEMTI framework includes plans to provide two different DTDs for taxons, one for each of the classifications, linked by the XPointer mechanism, and provision of a means to allow evaluators to parameterise the MT quality model depending on the context of use (currently already under development).

Future work by NIMMWG is planned in order to deepen and refine work on recommendations, as these are as yet at a preliminary stage, and to integrate results from both sides of the Atlantic. The recently started INTERA project is expected to lead to further refinement of IMDI guidelines, as is close cooperation of IMDI with OLAC. OLAC continues to set up new working groups and these will likely involve ISLE members. It is expected that there will be gradual merging of metadata-related proposals within the context of the Semantic Web and as an outcome of deliberations of ISO TC37/SC4. Work on annotation tools will be continued within the NITE project. In the US, further work on AGTK is planned, to cover video support and coding of gesture. As the US part of ISLE finishes later than the EU part, work will continue there for a short period, mainly to wrap up and document the annotated resources produced during the project for delivery to

the LDC. Both EU and US ISLE members will in the near term pursue support for the creation of a large discourse-tagged corpus for task-oriented dialogues of various types.

All ISLE members will continue to promote the EAGLES initiative through publication, participation in international events and in national and international projects.

9. Deliverables List

Deliverable No.	Deliverable Title	Workpackage No.	Lead Participant	Estimated Person-months	Deliverable Type	Security ²	Delivery (Project Month)
D14.1	Consortium Agreement	14	CPR	0.5	O	Int.	9
D13.1	Chapter on Evaluation of MT Systems for the Evaluation Handbook: First Draft	13	ISSCO	2	R	Int.	10
D11.1	Survey of Existing Tools, Standards and User Needs for Annotation of Natural Interaction and Multimodal Data	11	SDU	7	R	Pub.	13
D10.1	Survey of Best Practice and Trends in Metadata Description	10	SDU	13	R	Pub.	14
D2.1	Extension of EAGLES Guidelines on Lexicon/Semantics: Survey Part	2	CPR	6	R	Pub.	14
D3.1	Survey of Major Approaches to Bilingual/Multilingual Lexicons	3	CPR	7	R	Pub.	14
D11.2	Requirements Specification for a Tool in Support of Annotation of Natural Interaction and Multimodal Data	11	SDU	7	R	Pub.	19
D11.3	Demonstration of NIMM Implementation of Tool Elements to Illustrate Proof of Concept	11	SDU	7	P	Int.	19

² Int. = Internal circulation within project (and Commission Project Officer, if requested).
 Rest. = Restricted circulation list (specify in footnote) and Commission Project Officer only.
 IST = Circulation within IST Programme participants.
 FP5 = Circulation within Framework Programme participants.
 Pub. = Public document.

Deliverable No.	Deliverable Title	Workpackage No.	Lead Participant	Estimated Person-months	Deliverable Type	Security³	Delivery (Project Month)
D8.1	Survey of NIMM Data Resources including Strategic Report describing Current and Future User Profiles, Markets and User Needs for NIMM Resources	8	SDU	9.5	R	Pub.	21
D9.1	Survey of Annotation Schemes and Identification of Best Practice	9	SDU	16	R	Pub.	21
D10.2	Final Version of Proposed Standard for Meta-descriptions	10	MPI	13	R	Pub.	24
D10.3	Internet Site with Portal	10	MPI	6	R	Pub.	24
D4.1	Lexical Entry Tool (Software) and Accompanying Documentation	4	CPR	8	P	Pub.	26
D5.1	About 100 Lexical Entries for 2 or 3 Languages Cross-linked	5	CPR	10	O	Pub.	33
D5.2	Semantically Tagged Text for a Few Languages	5	CPR	10	O	Pub.	33
D13.2	Chapter on Evaluation of MT Systems for the Evaluation Handbook: Pre-final Draft	12	ISSCO	1.5	R	Pub	33
D2.2	Extension of EAGLES Guidelines on Lexicon/Semantics: Recommendation Part	2	CPR	8	R	Pub.	34
D3.2	EAGLES Guidelines for Bilingual/Multilingual Lexicons	3	CPR	11	R	Pub.	34
D6.1	Report with Evaluation Measures for Lexicons	6	CPR	7	R	Pub.	34
D6.2	Report with Results of Users' Evaluation	6	CPR	5	R	Pub.	34

³ Int. = Internal circulation within project (and Commission Project Officer, if requested).
 Rest. = Restricted circulation list (specify in footnote) and Commission Project Officer only.
 IST = Circulation within IST Programme participants.
 FP5 = Circulation within Framework Programme participants.
 Pub. = Public document.

Deliverable No.	Deliverable Title	Workpackage No.	Lead Participant	Estimated Person-months	Deliverable Type	Security⁴	Delivery (Project Month)
D1.1	Introduction to the Final CLWG Guidelines	1	CPR	1	R	Pub	36
D7.1	Introduction to and Summary of the Final NIMMWG Guidelines	7	SDU	8	R	Pub.	36
D8.2	Guidelines for the Creation of NIMM Data Resources	8	SDU	13	R	Pub.	36
D9.2	Guidelines for Multilayer Annotation of Unimodal and Multimodal Content	9	SDU	35	R	Pub.	36
D14.2	Final Report: Introduction to the EAGLES Guidelines; CD-ROM with Final Project Results	13	CPR	1	R	Pub.	36

⁴ Int. = Internal circulation within project (and Commission Project Officer, if requested).
 Rest. = Restricted circulation list (specify in footnote) and Commission Project Officer only.
 IST = Circulation within IST Programme participants.
 FP5 = Circulation within Framework Programme participants.
 Pub. = Public document.