

ATLAS*:

Traduzione automatica nella Lingua dei Segni

Alberto **Morello**
Rai - Centro Ricerche e Innovazione Tecnologica

Paolo **Prinetto**
Politecnico di Torino - Dip. di Automatica e Informatica

1. INTRODUZIONE

L'inserimento nel mondo della comunicazione elettronica è un passo importante sia per migliorare l'accesso delle persone alla società dell'informazione sia per superare le esclusioni sociali.

Le difficoltà di integrazione delle persone che sono nati sorde, o sono diventate sorde nei primi anni di vita, sono maggiori perché non hanno avuto la possibilità di acquisire la conoscenza della lingua parlata: più della metà delle persone sorde incontrano difficoltà nel leggere la forma scritta di una lingua parlata e, di conseguenza, hanno minori possibilità nell'apprendimento. D'altro canto, solo una minoranza fra i sordi è in grado di leggere le labbra abbastanza bene per capire la lingua parlata. Quindi, le persone non udenti hanno difficoltà a conversare con coloro che non utilizzano la lingua dei segni, a guardare film e televisione, a leggere.

Sommario

Il progetto ATLAS ha lo scopo di realizzare un sistema per tradurre dalla lingua italiana scritta alla Lingua dei Segni Italiana (LIS) e di visualizzare il risultato direttamente su supporti diversi attraverso personaggi virtuali. Il sistema, facendo ricorso ad approcci sia statistici sia basati su regole, traduce prima un testo italiano in un linguaggio intermedio, che convoglia sia elementi di informazione linguistica sia di visualizzazione. Rappresentazioni intermedie vengono poi utilizzate da un sistema di pianificazione per generare sia lo scenario in cui saranno eseguiti i segni sia i comandi relativi motore grafico. Il servizio è destinato alla distribuzione mediante diverse piattaforme, tra cui DVB, mobile, DVD, e web. L'articolo presenta l'organizzazione complessiva del sistema ATLAS, l'architettura e le basi di conoscenza.

* Questo articolo è basato sulla traduzione del documento "ATLAS: Automatic Translation into Sign Languages" presentato alla Conferenza IBC 2010 - Amsterdam, 9-14 Settembre 2010. Un precedente articolo sul progetto ATLAS è stato pubblicato nel numero di agosto 2009 di questa rivista.

La lingua dei segni (SL) è un linguaggio visivo che, per comunicare, si basa sulla gestualità del corpo invece di basarsi sui suoni. Si basa sulla combinazione contemporanea di forme, orientamenti e movimenti delle mani e di componenti non manuali, quali le espressioni facciali. Una lingua dei segni non è una pantomima, né è la resa visiva del relativo linguaggio verbale.

Le lingue dei segni si sviluppano ovunque esistano le comunità di non udenti, comunità che possono includere, oltre a persone sorde o con udito seriamente compromesso, anche interpreti, amici e famiglie di persone sorde.

Le lingue dei segni permettono ai bambini sordi di acquisire uno pieno sviluppo cognitivo all'interno della loro comunità, composta da persone udenti e non udenti. Tale sviluppo cognitivo rappresenta il presupposto di base a un pieno accesso all'istruzione, alla cultura, e all'inserimento nel l'ambiente sociale e lavorativo.

La "Sintesi dei Segni" è spesso vista come una soluzione al problema di comunicazione che ricorre spesso tra le persone non udenti e il resto della comunità, da qui nasce l'ipotesi che per risolvere il problema della comunicazione sia sufficiente convertire la voce o la scrittura in segno. La difficoltà di questo approccio nasce dal fatto che ogni lingua dei segni è molto diversa da tutte le lingue parlate e somiglia molto poco alle lingue più diffuse parlate nello stesso paese. Pertanto la traduzione da una lingua parlata alla lingua dei segni è un'impresa complessa. Alcune applicazioni di sintesi dei segni cercano di evitare questo problema usando versioni basate sui segni delle lingue parlate, ma questi sistemi non producono risultati che non sono comprensibili dalle persone sorde molto di più di quanto lo siano le versioni scritte di tali lingue.

I settori in cui sintesi offre i risultati più promettenti sono quelli relativi ai modi con cui registrare e rappresentare le lingue dei segni. Il modo più comune di registrare le lingue parlate è la scrittura, ma le lingue dei segni non sono quasi mai scritte. Ci

sono stati vari tentativi di creare sistemi di scrittura per le lingue dei segni, ma nessuna comunità di sordi ne ha adottato una in modo significativo. Una soluzione a questo problema che trova sostenitori è l'uso di sequenze video, ma questo approccio pone ulteriori problemi legati all'elevata capacità richiesti per la memorizzazione o la trasmissione dell'informazione video.

La Sintesi dei Segni è in grado di risolvere questi problemi: una frase da memorizzare può essere scritta, dall'autore o da una terza parte, usando uno dei sistemi di scrittura esistenti; può essere quindi memorizzata e trasmessa in forma scritta, per essere successivamente utilizzata per sintesi in un insieme fluente di segni, ogni qual volta sia necessario riprodurla.

L'unico modo per convertire a costi contenuti da una lingua parlata alla lingua dei segni è la traduzione automatica. Purtroppo, i linguisti hanno tentato per decenni di ottenere la traduzione automatica tra lingue dell'Europa occidentale strettamente correlate e hanno ottenuto un successo limitato. Le prospettive sono molto meno promettenti nel caso di una coppia di lingue così diverse come la lingua italiana scritta e parlata e la "Lingua dei Segni Italiana" (LIS), la lingua dei segni utilizzato in Italia.

L'obiettivo del "Riconoscimento della Lingua dei Segni" (SLR) è quello di fornire un meccanismo efficiente e accurato per trascrivere in testo la lingua dei segni. In letteratura, molti ricercatori si concentrano principalmente sul riconoscimento delle lingue dei segni limitando il vocabolario, a dimensioni piccole o medie. Le principali sfide che SLR deve ora affrontare è lo sviluppo di metodi che risolvano i problemi legati ad un ampio vocabolario e alla riproduzione mediante una sequenza continua di segni.

La Lingua dei Segni Italiana (LIS) è un lingua autonoma, utilizzata dai sordi per i quali è sia la lingua madre sia la lingua principale e dai quali l'italiano è percepita come una lingua straniera. La sottotitolazione di un programma televisivo, ad esempio, non può offrire la stessa profondità di informazione che

è invece garantita dalla lingua dei segni, caratterizzata dalla capacità di facilitare la partecipazione e la comprensione.

Le organizzazioni di servizio pubblico hanno il dovere di estendere i servizi a tutta la comunità, comprese le minoranze e le persone con disabilità sensoriali. Attualmente, in Italia, i programmi televisivi che offrono la traduzione simultanea nella lingua dei segni sono per lo più diffusi in orari di minor ascolto poiché, sottraendo una parte significativa dello schermo per mostrare l'interprete, incontrano un minor gradimento da parte della generalità del pubblico. E' quindi molto sentita l'esigenza di offrire la traduzione in LIS come servizio opzionale, selezionabile sulla TV digitale. Si prevede una crescente richiesta di interpretazione LIS nei contesti educativi, legali e sanitari e, nel prossimo futuro, in quelli relativi alla cultura e al divertimento.

Inoltre, non esistono oggi sistemi di traduzione completamente automatizzati dal testo alla lingua dei segni con visualizzazione mediante un personaggio virtuale che siano ampiamente adottati ed efficaci. Come dimostrato dai risultati di alcuni progetti internazionali, una soluzione basata sull'uso di personaggi virtuali è generalmente accettata dalla comunità dei non udenti ed in particolare i giovani sono molto interessati a questa tecnologia, in quanto offre loro una nuova possibilità di inserimento nella comunità globale.

Per quanto riguarda il punto di vista scientifico, una catena completamente automatica di traduzione da testo a LIS presentata da un personaggio virtuale rappresenta un argomento di ricerca del tutto nuovo e stimolante, che richiede la convergenza di diverse competenze specialistiche (traduzione automatica, ontologie, animazione in tempo reale).

Lo scenario descritto rende evidente la rilevanza della disponibilità di un sistema automatico a basso costo per coprire, in modo sostenibile, la richiesta crescente di una vasta gamma di servizi con contenuti LIS.

2. IL PROGETTO ATLAS

ATLAS (www.atlas.polito.it) è un progetto triennale (gennaio 2009 - dicembre 2011), co-finanziato dalla Regione Piemonte nell'ambito del quadro "Tecnologie convergenti - CIPE 2007" (Settore di Ricerca: Scienza Cognitiva e ICT).

Il progetto mira a sviluppare strumenti a supporto della traduzione dalla lingua scritta italiana alla lingua dei segni italiana. In particolare, l'obiettivo concreto del progetto è lo sviluppo di un traduttore in grado di tradurre frasi dalla lingua italiana naturale nella corrispondente sequenza di segni, attraverso una traduzione intermedia in forma scritta della LIS. Ciò comporta l'analisi del linguaggio naturale sia dal punto di vista linguistico sia da quello strutturale e la realizzazione di un interprete virtuale.

Il progetto mira a raggiungere i seguenti obiettivi::

- Sviluppare un sistema di traduzione automatica da un testo scritto in lingua italiana (sottotitoli o testi in generale) a LIS. La traduzione sarà realizzata ricorrendo alle tecniche più avanzate basate sulla statistica e sulle regole.
- Sviluppare *corpora* di informazioni appropriato per addestrare correttamente il sistema di traduzione automatica. Ci si focalizzerà su ambiti specifici, tra cui le informazioni meteo, le guide artistiche o territoriali, e su ulteriori argomenti individuati nel corso del progetto.
- Sviluppare un editor visuale assistito per supportare gli interpreti nella traduzione da testi in italiano a LIS. L'editor servirà come strumento di riferimento per la costruzione del *corpus*.
- Sviluppare un flusso per il *rendering* e la visualizzazione dell'interprete virtuale, di seguito denominato "personaggio virtuale".
- Definire i comandi per il personaggio virtuale per le diverse piattaforme di distribuzione e i vari dispositivi *client*.
- Progettazione e realizzazione di piattaforme *client-server* per dimostrare il servizio su diverse piattaforme, incluse i diversi tipi di Digital Video Broadcasting (DVB), telefoni cellulari, applicazioni web, e ulteriori supporti fisici come i DVD.

Fig. 1 - L'Architettura ATLAS.

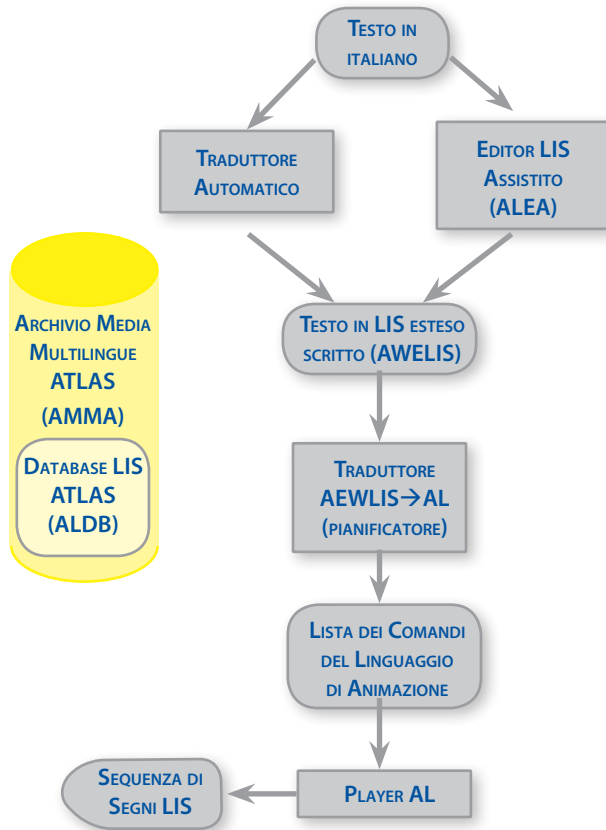
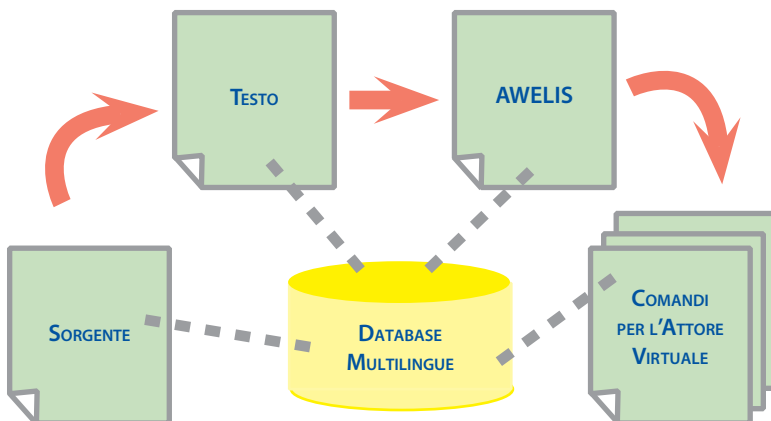


Fig. 2 - Catena di traduzione



3. L'ARCHITETTURA ATLAS

L'architettura del sistema ATLAS è mostrata in figura 1.

I testi scritti in linguaggio naturale italiano sono dapprima tradotti in una rappresentazione intermedia, denominata AEWLIS (ATLAS Extended Written Italian Sign Language), definita e sviluppata nella parte iniziale del progetto.

Le frasi espresse in AEWLIS sono poi tradotte in nella lingua gestuale del personaggio AL (Animation Language) che rappresentano la traduzione in LIS delle frasi stesse.

Più precisamente, il processo di traduzione dalla lingua italiana in LIS è suddivisa in diverse fasi principali (figura 2):

1. acquisizione del contenuto
2. traduzione del testo in AEWLIS
3. generazione dei comandi per l'Attore Virtuale
4. rendering dei comandi.

3.1 ACQUISIZIONE DEI CONTENUTI

L'acquisizione del contenuto sorgente è la prima fase del processo di traduzione e mira a raccogliere diversi tipi di contenuto: audio, video, pagine di televideo, pagine web, testi e sottotitoli. I componenti multimediali sono importanti anche per la generazione di un servizio completo su canali di comunicazione quali la televisione digitale o i dispositivi mobili.

3.2 L'AEWLIS

Nella fase di ideazione del progetto ATLAS, i proponenti ritennero che uno dei compiti principali del progetto doveva riguardare la definizione di una forma scritta della Lingua dei Segni Italiana.

Questa ipotesi è basata su diverse motivazioni:

- La maggior parte delle lingue (parlate) hanno, come uno dei principali modi di espressione, una forma scritta.
- Molti sforzi sono stati fatti per definire i modi per rappresentare in forma stampata la sequenza di segni di una "frase". Questo è stato fatto per la maggior parte delle lingue dei segni del mondo, con l'intendimento che la notazione scritta deve essere il più possibile indipendente dalla lingua.
- Una specie di rappresentazione "simbolica" del flusso dei segni è necessaria per attuare correttamente il processo di traduzione; in particolare, per fornire di un flusso di input adatto il modulo responsabile della animazione dei personaggi.
- Tale rappresentazione è necessaria anche per lo sviluppo di un corpus "annotato" dei "testi" segnati.

Si è tenuto conto in questo modo delle peculiarità del LIS (e del linguaggio dei segni in generale) nella definizione del AEWLIS. Tra queste, una delle più rilevanti è la simultaneità di segni: i suoi componenti costituenti, il "cheremes" (la forma di mano, e l'orientamento e il movimento di mani, braccia e corpo) si verificano contemporaneamente; al contrario, i componenti delle lingue parlate e scritte (i "fonemi" e i "grafemi", rispettivamente) sono per lo più di natura sequenziale. Le espressioni facciali sono inoltre di grande importanza per attribuire un significato. Inoltre, tutti i SL stabiliscono "loci" di riferimento nello spazio, cioè luoghi speciali nello spazio dei segni che rappresentano persone, luoghi, e così via.

In particolare, AEWLIS eredita le specifiche morfologiche e la struttura sintattica della LIS; i seguenti canali di comunicazione sono stati considerati rilevanti in LIS: Mani, Direzione, Corpo, Spalla, Testa, Viso, Labiali, e Sguardo. I canali sono ritenuti ortogonali tra loro, si richiede una registrazione indipendente delle azioni eseguite su ciascuno di essi.

L'insieme dei segni utilizzati nell'ambito del progetto ATLAS include:

- Segni ampiamente utilizzati nell'ambito delle varie comunità italiane di sordi.

- Segni nuovi definiti nell'ambito del progetto ATLAS, precedentemente non esistenti, ma necessari per rappresentare concetti specifici, in particolare associati a eventi e situazioni meteorologici.
- Segni associati con "classificatori LIS".

Nell'ambito dell'archivio multimediale ATLAS (AMMA) è stato definito e utilizzato un database per memorizzare tutti gli elementi di informazione associata a ciascuno dei segni ATLAS.

3.3 TRADUZIONE IN AEWLIS

La traduzione da italiano scritto in testo AEWLIS è la parte più difficile del processo generale e si basa su algoritmi statistici e matematici molto complessi.

Ci sono fondamentalmente due approcci per adempiere a questo compito: traduzione basata sulle regole e traduzione basata sulla statistica.

Il primo approccio, attualmente adottato dal team del Dipartimento di Informatica della Università di Torino, è basata su regole per la mappatura della struttura grammaticale e sintattica della lingua di partenza sulla lingua di arrivo (AEWLIS). L'analisi comprende quattro fasi principali: (1) analisi morfologica e dizionario di accesso (2) analisi sintattica (3) interpretazione semantica (4) traduzione in notazione ATLAS (Trame).

Il secondo approccio, adottato dal gruppo di ricerca della Fondazione Bruno Kessler di Trento, si basa su metodi statistici, e si basa su algoritmi classici di apprendimento automatico.

Si è seguito questo doppio percorso al fine di individuare i vantaggi e gli svantaggi dei due metodi, e di operare in conclusione la scelta definitiva, che potrebbe consistere nella integrazione dei due approcci.

3.4 ANIMAZIONE DELL'INTERPRETE VIRTUALE

La traduzione della rappresentazione AEWLIS nella personaggio AL si realizza in due fasi. Il primo passo consiste nella traduzione di espressioni AEWLIS in AL, mentre il secondo passo si occupa dell'analisi



Fig. 3 - Uno dei personaggi virtuali ATLAS.

della sequenza AL al fine di produrre le animazioni, cioè della vera e propria riproduzione animata.

D'altra parte, la AL include componenti specifiche volte a descrivere come produrre e collegare i movimenti di base al fine di generare una sequenza di animazioni comprensibili da far attuare al personaggio virtuale. La figura 3 mostra uno dei personaggi virtuali sviluppati, nell'ambito del progetto, dal Virtual Reality & Multimedia Park di Torino.

Sono due i livelli da considerare nell'animazione: il "livello del segno", dove i segni LIS sono gli elementi fondamentali e le regole limitano la loro concatenazione in una frase LIS, e il "livello gestuale", dove i segni sono oggetti complessi costituiti di gesti base del personaggio e le regole che vincolano la loro combinazione sequenziale e parallela e per ottenere i segni ammessi. Il livello gestuale si occupa dei movimenti relativi ai segni elementari che riguardano mani, braccia, spalle, collo, testa e tronco, in combinazione con le espressioni facciali. Accade che le modalità diverse di un unico gesto hanno significati diversi in lingua LIS, e pertanto un aspetto importante di questo studio è tener conto della parametrizzazione dei gesti di base, così come le regole di combinazione, per velocità, accelerazione, larghezza e lunghezza. Lo studio a livello gestuale deve comprendere quali segni saranno da considerare nel loro insieme (le entità

atomiche) e quali saranno i segni a dover essere decomposti in gesti (le entità subatomiche) per poi essere ricomposte in tempo reale.

Le espressioni facciali e labiali sono ampiamente utilizzati dagli interpreti umani e sono una parte significativa della comunicazione. Un buon esempio può essere offerto dalla forma interrogativa: nel linguaggio LIS la struttura della frase non cambia, come, per esempio, in inglese, e una domanda è rappresentata da una frase affermativa accompagnata da un'espressione interrogativa del viso.

I personaggi ATLAS sono progettati per offrire un controllo totale di espressività: modellazione dei muscoli, simulazione della pelle, modellazione antropometrica e sincronizzazione del parlato consentono al personaggio virtuale di fornire traduzioni molto espressive e realistiche. I "canali di comunicazione" forniti dal AEWLIS sono utilizzati per controllare l'interprete virtuale.

Il processo di traduzione utilizza metodi differenti per trasferire l'espressione giusta all'interprete virtuale: ad esempio, quando vi è un punto interrogativo nella frase da tradurre, l'interprete virtuale assume una espressione facciale interrogativa.

Nella traduzione basata su regole, l'algoritmo di analisi semantica genera tutti i dati relativi alle espressioni.

Nella traduzione basata sulla statistica, nel corso del lavoro di apprendimento del motore statistico ogni segno acquisito è associato ai dati relativi, comprese la posizione di mani, direzione, corpo, spalle, testa, viso, labbra, sguardo e espressioni. L'uscita della traduzione è una sequenza di segni in cui vengono introdotte le informazioni di cui sopra, incluse le espressioni facciali, sulla base di considerazioni statistiche.

Le sequenze risultanti AL sono rese per mezzo di un motore di *rendering* 3D. Un sistema automatico di "fusione di animazione" farà in modo che le animazioni generate LIS includano transizioni uniformi tra segni. Infine, il motore sarà distribuito a un numero limitato di utenti finali e i loro *feedback* saranno raccolti e analizzati sotto la supervisione di un esperto.

Gli strumenti sviluppati saranno applicabili a diversi domini, tra cui la traduzione simultanea in LIS dei sottotitoli associati a programmi TV, la presentazione di contenuti digitali e interfacce per l'accesso del cittadino ai servizi pubblici, la presentazione di informazioni e di messaggi diretti alle persone sorde su dispositivi mobili utilizzando il LIS, creazione di un canale televisivo dedicato ai sordi su base nazionale o regionale. .

3.5 IL DATABASE MULTIMEDIA DI ATLAS

E' stato realizzato un database multimediale, denominato AMMA, per supportare il processo di traduzione sopra citato (figura 4). Sono necessari diversi componenti, compresi i database e i *repository* di file multimediali, per memorizzare il set complessivo degli elementi di informazione gestiti all'interno del processo di ATLAS.

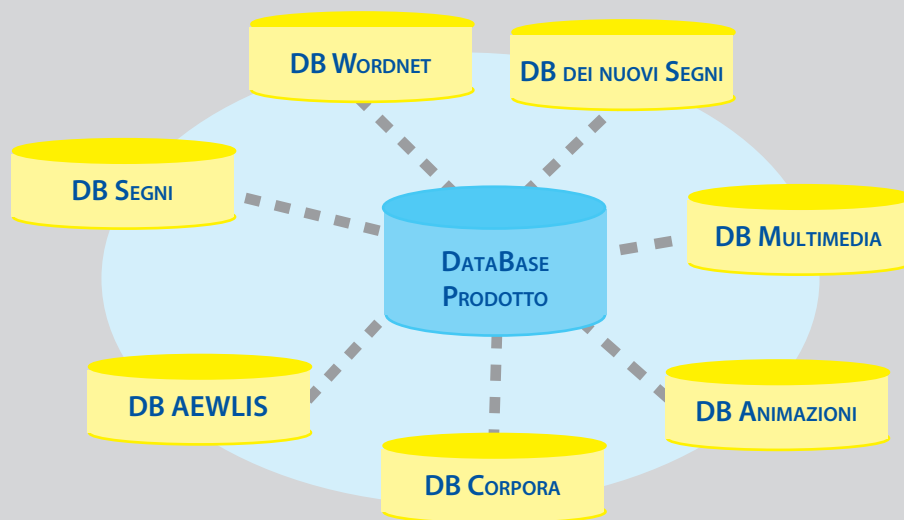
3.6 LA RETE ATLAS

E' stata realizzata una rete completa per consentire una corretta interconnessione di tutti gli utenti ATLAS (figura 5).

Gli utenti ATLAS sono classificati sulla base del loro ruolo, distinguendo tra:

- **i partner ATLAS:** le istituzioni e centri di ricerca coinvolti nel progetto. Essi includono: Politecnico di Torino (Dip. di Automatica e Informatica), Università di Torino (Dip. di Informatica e Dip di Psicologia.), la Rai Radiotelevisione Italiana SpA, BEPS Engineering, Microsoft Innovation Center, Virtual Reality and Multimedia Park, Lumiq Studios srl, Alto Sistemi srl, CSP - Innovazione nelle

Fig. 4 - Organizzazione del database AMMA.



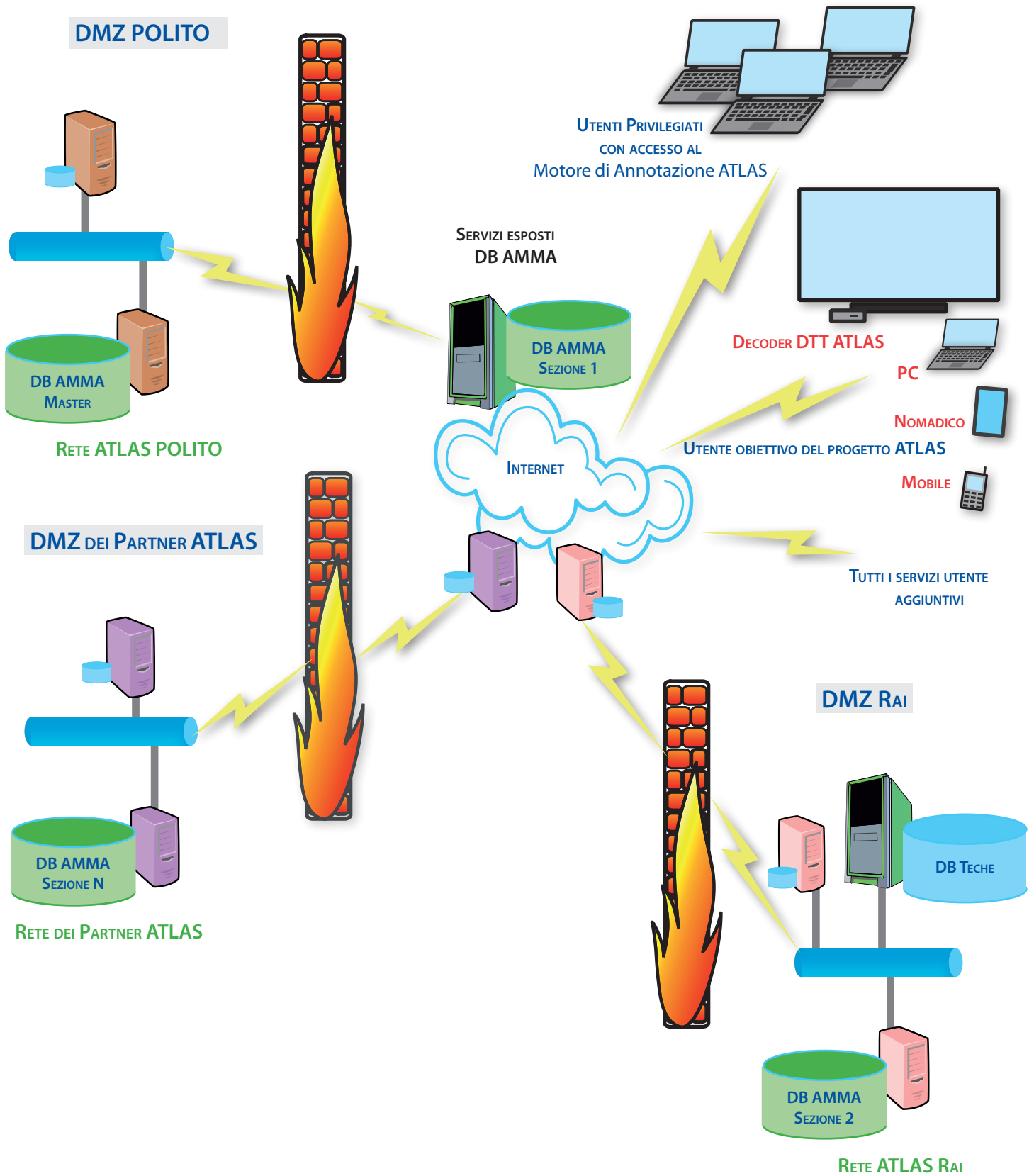


Fig. 5 - La rete ATLAS. Le DMZ (*demilitarized zone*) sono segmenti isolati di LAN (una "sottorete") raggiungibili sia da reti interne che esterne che permettono, però, connessioni esclusivamente verso l'esterno: gli *host* attestati sulla DMZ non possono connettersi alla rete aziendale interna.

ICT, Fondazione Bruno Kessler - Trento, Cooperativa SCS Global Communication, University of Illinois a Chicago.

- **ATLAS Power Users:** utenti specializzati incaricati della creazione di corpora necessari per sostenere e migliorare la fase di traduzione statistica.
- **ATLAS Target Users:** le persone sorde che beneficeranno della disponibilità degli interpreti ATLAS virtuali sui mezzi di comunicazione, tra cui TV digitale, inter-net, DVD, smart-phone, ...
- **Tutti gli utenti** che possono accedere, attraverso il portale ATLAS, a tutti i prodotti e ai prodotti derivati sviluppati nell'ambito del progetto. Essi includono, tra gli altri, un completo dizionario LIS per smart-phone, un potente editor per assistere l'annotazione dei film in AEWLIS, il porting del motore grafico *open-source* 3D OGRE (*Object-Oriented Graphics Rendering Engine*) per piattaforme Windows CE ... In conclusione, è bene ricordare che tutti i programmi, i dati e corpora sviluppati nell'ambito del progetto ATLAS saranno messi a disposizione delle varie comunità di ricerca mediante licenze *open-source*.

3.7 VALUTAZIONE DELLE TRADUZIONI E COMPRESIBILITÀ DELL'INTERPRETE VIRTUALE

Uno specifico *Work Package* del progetto è dedicato a valutare l'efficacia complessiva del processo.

In particolare, un primo passo consiste nella valutazione della traduzione da italiano a AEWLIS. E' condotto da interprete madrelingua LIS e utenti sordi, che partecipano alla preparazione del materiale per il processo di apprendimento del traduttore statistico. Al momento di questo documento la fase di apprendimento è ancora in corso: le valutazioni effettuate sulle frasi di esempio sono considerati positivi.

Un secondo passo consisterà nella valutazione sia dell'animazione sia dei componenti di traduzione: queste attività saranno svolte attraverso esperimenti condotti con persone sorde per valutare la qualità dei risultati.

Il progetto è sostenuto dalla collaborazione di ENS (Ente nazionale sordomuti), l'Associazione Nazionale Italiana sordo persone, che svolge un ruolo importante nel processo di validazione.

BIBLIOGRAFIA

- Merlo Pick V. (1988), "Grammatica della Lingua Swahili", E.M.I., Torino.
- Radutzky E. (1992), "Dizionario bilingue elementare della Lingua dei Segni Italiana", ed. Kappa, Roma.
- Romeo O. (1996), "Dizionario dei Segni", Zanichelli, Bologna
- Smith M. (2003), "On the Interpretation of Restrictive Internally Headed Relative Clauses".
- Vaijayanthi Sarma (2003), "Non-canonical word order" in "Word Order and Scrambling" a cura di Simin Karimi, 2003, Blackwell Publishing, UK.
- Franchi M. L. (1987), "Componenti non manuali" in "La lingua italiana dei segni" a cura di V. Volterra, Il Mulino, Bologna.
- Holmes P. and Hinchliffe I. (1997), "Swedish – An essential Grammar", Routledge.
- Laudanna A. (1987), "Ordine dei segni nella frase" in "La lingua italiana dei segni", V. Volterra, Il Mulino, Bologna.
- Verdirosi M. L. (1987), "Luoghi" in "La lingua italiana dei segni" a cura di V. Volterra, 1987, Il Mulino, Bologna.
- Vries, M. de (2002), "The Syntax of Relativization" PhD Dissertation at University of Amsterdam, published by LOT, Utrecht.

RICONOSCIMENTI

Gli autori desiderano ringraziare tutte le persone coinvolte nel progetto ATLAS per la fornitura del materiale e degli elementi di informazione necessari per realizzare questo articolo.