

Il processo di codifica dei dati testuali dell'indagine Multiscopo "Uso del tempo"

Maria Clelia Romano, Tania Cappadozzi¹

ISTAT – SDS/2 – Via Ravà 150 – 00142, Roma – Italy
romano@istat.it

Abstract

During the performance of Time Use surveys the climax is the coding phase of daily activities performed by respondents and freely expressed in the diary. In this paper will be described the main choices adopted by Italian National Institute for Statistics during the planning of Time Use Survey 2002-2003, specially with reference to the textual data collected treatment. Two important changes have been introduced: first, the data entry of textual data reported in the diaries, that allowed to build a very innovating database for dimension and content; second, the implementation of an assisted coding system that allowed to monitor whole coding process and to better the quality of final data.

Riassunto

Una fase cruciale nella realizzazione delle indagini sull'uso del tempo è rappresentata dalla codifica delle attività giornaliere svolte dagli intervistati ed espresse in forma libera nei diari. Nel presente lavoro descriveremo le principali scelte adottate dall'Istituto Nazionale di Statistica nel corso della riprogettazione dell'indagine 2002-2003, con riferimento al trattamento dei dati testuali raccolti. In particolare, due importanti novità sono state introdotte in questa fase: la registrazione dei dati testuali riportati nei diari, che ha condotto alla costruzione di un archivio dati per dimensioni e contenuto di grande rilevanza, e l'implementazione di un sistema di codifica assistita che ha consentito di monitorare l'intero processo di codifica e di migliorare la qualità del dato finale.

Keywords: time use, textual data, daily activities, hierarchical classification, coding list adaptations.

1. L'indagine Istat Multiscopo "Uso del tempo" 2002-03

Tra aprile 2002 e marzo 2003 si è svolta la rilevazione Istat Multiscopo sulle famiglie "Uso del tempo" 2002-03. Nel corso di tale indagine è stato intervistato un campione di 21.075 famiglie per un totale di circa 55.000 individui.

Gli strumenti di rilevazione utilizzati sono quattro: un questionario per intervista individuale, un questionario per intervista familiare, un diario settimanale destinato alle persone di 15 anni e più mirato a rilevare le ore di lavoro retribuito svolte nel corso di una data settimana, e un diario giornaliero destinato alle persone di 3 anni e più e destinato a rilevare specificamente l'uso del tempo quotidiano.

Lo strumento principe di questa indagine è ovviamente rappresentato dal diario giornaliero, che costituisce una fonte preziosa di informazioni ad elevato livello di dettaglio sull'organizzazione dei tempi individuali, ed apre, essendo l'unità di rilevazione la famiglia, la possi-

¹ Il lavoro è frutto della collaborazione delle due autrici. Tuttavia la stesura dei paragrafi 1, 2, 3, 7 è da attribuire a Maria Clelia Romano e quella dei paragrafi 4, 5, 6 a Tania Cappadozzi.

bilità di un'analisi interindividuale dell'uso del tempo, poiché consente di rilevare le varie combinazioni tra i tempi di tutti i componenti della famiglia.

Infatti, attraverso la compilazione del diario è possibile conoscere il modo in cui ciascun rispondente ripartisce 24 ore (divise in 144 intervalli di 10 minuti) tra le varie attività giornaliere, gli spostamenti, i luoghi frequentati e le persone con cui le ha trascorse (vedi fig. 1). Si tratta cioè di un'informazione che presenta un livello di dettaglio estremamente elevato, non comparabile con quelle ricavate dai tradizionali questionari a domande fisse.

L'indagine "Uso del tempo" rientra inoltre in un progetto Eurostat di armonizzazione delle indagini "Time Use" che nato nel 1994 è giunto alla formulazione di Linee Guida (Eurostat, 1999), che se rispettate nella progettazione delle indagini nazionali, consentono di perseguire l'obiettivo della comparazione internazionale dei risultati. Secondo tali linee guida e secondo quanto realizzato nella maggior parte dei paesi europei, l'uso del tempo rilevato attraverso lo schema del diario giornaliero viene poi classificato secondo un analitico elenco di codici numerici.

La descrizione delle attività e dei luoghi, dunque, pur essendo lasciata alla libera capacità di espressione di ciascun rispondente viene in seguito sottoposta ad una classificazione. In Italia, il sistema di classificazione adottato, pur partendo da quello suggerito da Eurostat, e conservando pertanto le conseguenti possibilità comparative, è stato integrato ed adattato alle specifiche esigenze nazionali.

	Che cosa sta facendo? Indichi le attività che svolge ad intervalli di tempo di 10 minuti!	Che cos'altro sta facendo? Indichi l'attività contemporanea più importante	Dove si trova o come si sta spostando?	E da solo o con persone che conosce?			
				Da solo	Con familiari conviventi meno di 10 mesi	Con familiari non conviventi	Con altre persone che conosce
07.00 - 07.10	Mi sono svegliata		A casa mia			X	
07.10 - 07.20	Mi sono lavata		↓	X			
07.20 - 07.30	Ho svegliato mio figlio				X		
07.30 - 07.40	Ho preparato la colazione	Ho ascoltato la radio		↓			
07.40 - 07.50	Ho fatto colazione	Ho letto il quotidiano		↓			
07.50 - 08.00	Ho lavato i piatti	Ho ascoltato la radio		↓			
08.00 - 08.10	Ho rifatto i letti	" "	↓				
08.10 - 08.20	Ho portato il figlio al nido	Ho parlato con figlio	A piedi			X	
08.20 - 08.30	Sono andata al lavoro		In autobus	X			
08.30 - 08.40	Lavoro (unica occupazione)		In ufficio				
08.40 - 08.50			" "				
08.50 - 09.00			" "				
09.00 - 09.10			" "				
09.10 - 09.20			" "				
09.20 - 09.30			" "				
09.30 - 09.40			" "				
09.40 - 09.50			" "				
09.50 - 10.00			" "				

Figura 1. Lo schema del diario giornaliero

Obiettivo di questo lavoro è descrivere le principali scelte effettuate dall'Istituto Nazionale di Statistica nella progettazione dell'indagine "Uso del tempo" 2002-03, con particolare riferimento alle strategie innovative adottate per la classificazione delle informazioni testuali rilevate tramite i diari giornalieri. In particolare, viene descritta l'esperienza della codifica assistita e dei principali vantaggi che da tale scelta sono derivati, in termini di possibilità di monitoraggio del processo di codifica e di miglioramento della qualità dei dati.

2. La codifica: caratteristiche e problematiche

Buona parte della riuscita di questa indagine dipende dall'accuratezza con cui viene svolto il lavoro di codifica delle informazioni contenute nei diari giornalieri. Solo l'eshaustività del sistema di classificazione e la corretta applicazione dello stesso possono garantire elevati standard di risultato, oltre che possibilità comparative a livello internazionale.

Il sistema di classificazione Eurostat ha una struttura gerarchica, fino a un massimo di quattro livelli, aperto alla possibilità di inserire codici specifici di interesse nazionale. Dieci sono le categorie di primo livello, individuate per classificare le attività in base agli obiettivi e al contesto (per esempio: Lavoro, Studio, Cura della casa e gestione della famiglia, Vita sociale etc.). Di seguito si riporta un esempio della struttura gerarchica:

3. CURA DELLA PROPRIA CASA E DELLA PROPRIA FAMIGLIA

3.1 CUCINARE, LAVARE E RIORDINARE LE STOVIGLIE

- 3.1.1 Preparare i pasti, le merende, la colazione
- 3.1.2 Preparare torte, biscotti, pane e altri alimenti con farina
- 3.1.3 Lavare le stoviglie
 - 3.1.3.1 Apparecchiare e sparecchiare la tavola
- 3.1.4 Preparare conserve, confetture, congelare, imbottigliare, ecc.
- .
- .
- . (*Altre categorie di terzo livello*)
- .
- 3.1.9 Altre attività specificate legate alla preparazione di alimenti

3.2 PULIZIA E RIORDINO DELLA CASA

- .
- . (*ALTRE CATEGORIE DI SECONDO LIVELLO*)
- .

3.9 CURA E AIUTI AD ADULTI DELLA FAMIGLIA

La raccolta dei dati sull'uso del tempo è finalizzata alla misurazione della durata delle attività, ma soprattutto ad una corretta caratterizzazione delle stesse in base al contesto, alle modalità con cui si succedono, agli scopi, al luogo e al momento del giorno, della settimana, della stagione in cui si svolgono. Proprio per la molteplicità delle dimensioni a cui la descrizione delle attività riportate sui diari rimanda, la fase di codifica è il "perno" della realizzazione di tale indagine. Essa rappresenta la chiave per estrapolare un codice sintetico dalle descrizioni riportate in forma testuale libera, con l'utilizzo di espressioni eterogenee e in funzione del bagaglio linguistico e della capacità espressiva di ciascuno (Camporese e Ranaldi, 2001).

Tuttavia, come noto, l'associazione testo-codice è tutt'altro che un'operazione semplice, quanto piuttosto il risultato di un processo: 1. le espressioni verbali devono essere comprese e

valutate nel loro contesto, che è fatto di informazioni sull'individuo in oggetto e sulle caratteristiche della specifica giornata; 2. Le informazioni ridondanti non devono essere considerate, ma vanno selezionate solo quelle utili; 3. Vanno dedotte le informazioni non scritte.

I codificatori sono dunque chiamati a fare una vera e propria traduzione da un linguaggio ad un altro (quello in codici) che si pone ad un livello di complessità, ricchezza ed eterogeneità molto lontano da quello originario. Un'altro elemento di divergenza tra i due linguaggi è il diverso grado di tolleranza per l'ambiguità. Espressioni vaghe, frasi dai molteplici significati sono frequentemente usate e riconosciute nel linguaggio verbale; al contrario, il linguaggio dei codici è rigido, non è in grado di gestire le ambiguità e la necessità di interpretazione.

Tuttavia, è proprio la capacità di contestualizzare l'attività la principale competenza necessaria all'attività di codifica: per scegliere correttamente un codice occorre concentrarsi sul diario come un tutto, immaginando l'ora del giorno, la stagione dell'anno, e tenendo ben presente le caratteristiche sociodemografiche del rispondente (Camporese e Ranaldi, 2001). In sintesi due aspetti devono essere distinti. L'azione svolta e il suo obiettivo, poiché due azioni eguali, possono essere codificate in modi differenti a seconda dei loro differenti obiettivi. Solo nel momento in cui si ha questa visione di insieme è possibile concentrarsi sulle singole parti del diario e individuarne gli episodi, interpretando nel giusto modo ciascuna attività descritta (per esempio, se cucinare significa cura della propria famiglia, aiuto informale ad altre famiglie, lavoro volontario, etc.)

3. Le principali novità nella progettazione del sistema di codifica dell'indagine 2002-2003: registrazione delle informazioni testuali e codifica assistita

Uno dei primi problemi che ci si è posti in fase di riprogettazione dell'indagine ha riguardato proprio le strategie migliori di applicazione della classificazione, in termini di strumenti tecnici e di personale a cui affidare questo delicato compito.

Nella precedente indagine svolta nel 1988 e nella pilota del 1996, l'esperienza della codifica ha fatto sorgere non pochi problemi ed ha richiesto l'individuazione di strategie alternative e soprattutto innovative rispetto al passato. Per la precisione, nel corso della precedente indagine "Uso del tempo" (1988) è stata adottata la codifica manuale, effettuata dai rilevatori sulla base di apposite liste di codici relativi a "tipo di attività", "luogo di svolgimento" e "persone presenti". Tale lavoro non sempre è stato condotto in maniera ottimale, al punto che si è reso necessario un impegnativo processo di revisione che quasi sempre ha comportato la rilettura dei diari originali, per verificare l'adeguatezza dei codici e adottare gli opportuni interventi correttivi. In pratica, i revisori, per eliminare tutti gli errori e le omissioni accumulate in sede di rilevazione, hanno effettuato un lavoro simile a quello già svolto dai codificatori, cioè la lettura sul materiale cartaceo dell'attività descritta per esteso, la ricerca del codice, il confronto con il codice registrato e l'eventuale correzione.

Anche nel corso dell'indagine pilota del 1996 il processo di codifica fu svolto manualmente, questa volta ad opera di codificatori appositamente addestrati. Pure questa esperienza ha rivelato molti limiti, soprattutto in assenza di un'adeguata e costante supervisione dell'attività di codifica.

La codifica manuale inoltre richiede un elevato investimento di risorse umane e di tempo (si tenga presente la dimensione campionaria), ed accresce i rischi connessi alla discrezionalità del codificatore, poiché limita le possibilità di monitoraggio e di interventi in corso d'opera, consentiti invece da un sistema di codifica informatizzato e gestito in rete.

La complessità del sistema di codifica adottato ed in particolare la necessità di interpretare e contestualizzare le attività, prima di procedere all'assegnazione del codice corretto hanno reso poco indicato anche il ricorso ad un software di codifica automatica. Tale software infatti si presta alla codifica di informazioni (come la professione, l'elenco dei comuni, etc.) per le quali è possibile l'attribuzione univoca di un codice. Il sistema di classificazione adottato cioè prevede un unico codice per ciascuna delle voci/espressioni che tale informazione può presentare ed è unica la variabile da tenere in considerazione. Non è questo il caso del sistema di codifica utilizzato nel corso dell'indagine "Uso del tempo", visto che identiche attività possono essere codificate in maniera diversa, in base all'orario in cui vengono svolte, alle finalità, ai destinatari, etc., dunque le variabili che concorrono alla individuazione del codice corretto sono molteplici.

La crucialità del processo di codifica nell'indagine "Uso del tempo" e l'esperienza accumulata nel corso delle precedenti indagini hanno indotto l'Istituto a ricercare strategie di codifica più efficaci di quelle già sperimentate. In particolare, l'obiettivo che ci si è posti è stato quello di ridurre fortemente la discrezionalità dei codificatori e dunque il rischio di commettere errori in questa delicata fase dell'indagine, cercando contemporaneamente di contenere i tempi di realizzazione del lavoro e di rendere possibile un efficace monitoraggio in corso d'opera.

Sulla base di tali considerazioni si è deciso di implementare un sistema di codifica assistita utilizzando il software Blaise. Tale strategia, pur fornendo uno strumento elettronico che facilita, velocizza e supporta l'attività umana, lascia comunque al centro dell'attività la capacità critica del codificatore.

La fattibilità di tale scelta dipendeva tuttavia da due condizioni: 1) la possibilità di prevedere la registrazione, per la prima volta nella storia di questa indagine, di tutte le informazioni testuali riportate nei diari e, dunque, la costruzione di un archivio testuale per contenuto e dimensione mai acquisito dall'Istituto Nazionale di Statistica, ma con elevati costi di acquisizione e gestione; e 2) la possibilità di disporre di informazioni testuali che consentissero di predisporre il dizionario necessario per implementare il sistema di codifica assistita.

Alla prima esigenza l'Istituto ha risposto positivamente ed ex post è possibile dire che l'investimento fatto in questa direzione rende giustizia al grosso sforzo fatto dalle famiglie italiane per rispondere all'impegno richiesto dalla compilazione dei diari e che risulta molto chiaro dalla lettura delle descrizioni riportate, in molti casi vere e proprie storie di vita giornaliera. Certo non mancano i diari compilati in maniera lacunosa, poco chiara, talvolta evidentemente svogliata; ciononostante sono la gran parte i diari in cui i rispondenti hanno descritto scrupolosamente e con un dettaglio a volte maniacale lo scorrere del tempo e l'avvicinarsi delle varie attività svolte. Insomma la ricchezza di informazioni contenute nei diari è veramente impressionante (vedi tab. 1). La stessa codifica, per quanto resa necessaria dall'esigenza di pervenire a una sintesi dei dati e consentire analisi comparative a livello internazionale, non poteva esaurire le esigenze conoscitive che pure questa indagine è in grado di soddisfare. Pertanto la registrazione, oltre che dei codici anche delle informazioni testuali contenute nei diari, rappresenta una decisione estremamente importante, poiché apre la strada ad un'infinità di analisi di tipo innovativo: si pensi non solo all'applicazione di tecniche di text mining ma anche agli studi qualitativi, che senza tale archivio informatizzato sarebbe stato quanto meno enormemente più complesso realizzare.

Per quanto riguarda la seconda condizione, si è deciso di recuperare i diari cartacei raccolti nel corso della pilota del 1996 e di predisporre la registrazione delle informazioni testuali in esso contenute. Sono stati pertanto registrati oltre 500 diari relativi a circa 200 famiglie.

Tale opportunità ha presto rivelato anche altre positive conseguenze, poiché il recupero e il trattamento di tali dati ha messo in luce alcune attività che per rilevanza sociale necessitavano di essere estrapolate dal codice generico nel quale ricadevano per essere individuate con codici specificamente ad esse attribuite (per es. il vedere cartoni animati rischiava di confondersi con il vedere genericamente la TV).

Tabola 1. Numero di stringhe raccolte tramite i diari giornalieri per tipo di informazione

	Attività principale	Attività contemp.	Luogo	Totale
N. stringhe	1.458.312	412.609	1.220.961	3.091.882
<i>di cui diverse</i>	240.480	58.057	31.769	312.221
N. occorrenze	4.991.028	1.477.765	2.909.917	9.378.710
<i>di cui diverse</i>	29.494	12.920	9.284	37.764

Fonte: Istat, Indagine Multiscopo "Uso del tempo" 2002-03.

4. Dalla ricchezza del testo all'integrazione del sistema di codici

Come accennato è stato necessario integrare il sistema di codici proposto da Eurostat con altri codici in grado di individuare attività socialmente rilevanti. Alcuni di questi erano stati già introdotti nel corso dell'indagine pilota 1996, altri sono stati introdotti ed utilizzati per la prima volta nell'indagine definitiva.

Per esempio, sono stati previsti codici più adatti alla classificazione delle attività dei bambini, la cui vita quotidiana e la cui percezione del tempo presentano delle grosse peculiarità rispetto alle esperienze degli adulti. Le loro attività hanno caratteristiche e specificità che le rendono più difficili da incasellare in un elenco di codici, pensato per sintetizzare principalmente le attività quotidiane degli adulti.

Infatti, il sistema originario di codici si presenta a tale proposito molto lacunoso, anche perché l'Italia è uno dei pochi Paesi al mondo a prevedere la compilazione dei diari per le persone al di sotto dei 10 anni (soglia minima di età suggerita da Eurostat).

I dati dell'indagine pilota sono stati una miniera preziosa di informazioni e di descrizioni indicative del modo di esprimersi dei bambini. Sulla loro base è stato pertanto stilato un elenco aggiuntivo di attività e di codici corrispondenti che è andato ad integrare il sistema di classificazione iniziale.

Tuttavia, oltre ad integrare il sistema di classificazione adottato con ulteriori codici attività, si è ritenuto importante alla luce delle informazioni testuali contenute nei diari dell'indagine pilota prevedere una serie di codici cosiddetti ausiliari (Istat, 2002a), che supportassero le successive analisi tematiche e di qualità, portando alla luce alcune caratteristiche tipiche della descrizione che gli individui fanno del loro tempo e che sarebbero andate perse nella traduzione in codici del linguaggio comune.

In particolare, alcuni di questi codici sono un prezioso supporto al complesso processo di controllo della qualità dei diari e di eventuali correzioni perché consentono di:

- individuare le contraddizioni tra le varie informazioni presenti nei diari (tra attività principale e contemporanea, tra attività e luogo, attività e persone presenti, etc.);
- individuare sequenze non ben descritte;

- individuare diari particolarmente lacunosi di informazione e con sequenze illogiche.

Altri codici ausiliari invece consentono di:

- individuare attività di durata chiaramente inferiore alla durata minima dell'intervallo (eventi puntuali);
- individuare la concentrazione di attività all'interno di uno stesso intervallo temporale (attività multiple);
- individuare informazioni aggiuntive fornite dai rispondenti utili a specifiche analisi di qualità e che sfuggirebbero alla normale attività di codifica.

Attraverso tali codici ausiliari potranno essere individuati più facilmente sia le attività più frequentemente connesse a certe particolari tipologie di compilazione dei diari, sia - cosa non meno interessante - le modalità linguistiche ad essi connesse.

Infine, non è stata dettagliata in codici, ma sarà un interessantissimo oggetto di analisi testuale invece la presenza di molteplici espressioni che, non rimandando ad una attività (ovvero ad un uso oggettivo del tempo), ricadono in un codice generico. Si tratta di espressioni che descrivono molto frequentemente le condizioni di contesto, uno stato d'animo, ma non un'azione. In altri termini pur non essendo stato chiesto agli intervistati di fornire tale tipo di dettaglio, in numerosi casi ciò è stato fatto. Sarà pertanto interessante cercare di capire quanto la necessità di connotare "emotivamente" l'uso oggettivo del tempo possa consentire l'identificazione di specifici gruppi sociali, o ancora meglio richiedere una riflessione metodologica in vista di una riprogettazione dell'indagine stessa.

5. Gli strumenti della codifica assistita

Concretamente la codifica delle informazioni testuali è avvenuta sfruttando le possibilità offerte dal software Blaise, che mette a disposizione un modulo per la codifica assistita basato su due tipologie di ricerca, l'albero dei codici (vedi fig.2) e il dizionario (vedi fig.3), che possono essere utilizzate singolarmente e/o congiuntamente (Istat, 2002b). Le due modalità di ricerca necessitano della preparazione di due diversi files: uno contenente il sistema dei codici della classificazione, a partire dal quale Blaise organizza una struttura gerarchica ad albero e, l'altro, un elenco delle frasi-attività associate ai codici corrispondenti.

Con la registrazione delle informazioni testuali (attività principale, attività contemporanea) dei diari dell'indagine pilota si è ottenuto un archivio di frasi-attività, sulla cui base è stato costruito il dizionario da fornire al software come ulteriore e privilegiato ambito linguistico al cui interno effettuare la ricerca del codice corretto.

In effetti, mediante il dizionario si cerca di avvicinare la classificazione al linguaggio comune, utilizzato dai rispondenti e di offrire ai codificatori uno strumento che, oltre a facilitare e velocizzare l'intero processo di codifica, riduca le incertezze di attribuzione dei codici e conseguentemente la discrezionalità del loro operato.

La ricerca sul dizionario della stringa di un'attività in lettura viene effettuata dal software Blaise mediante la sua scomposizione in sottogruppi di tre lettere (detti trigrammi) e il successivo confronto tra questo insieme di trigrammi e quelli presenti nel dizionario, in cui tutte le stringhe sono state preventivamente trasformate in altrettanti insiemi di trigrammi. Il risultato di tale ricerca consiste nel mostrare al codificatore le stringhe del dizionario con il maggior numero di trigrammi in comune con la stringa in lettura (vedi fig. 3).

Questo tipo di ricerca, da un lato, ha consentito di ridurre il numero di stringhe da inserire nel dizionario; poiché, ad esempio, inserendo nel dizionario la stringa “dormire” tutte le stringhe che contengono le tre lettere “dor” (come ho dormito, sto dormendo, dormivo ecc.) avrebbero fatto riferimento ad essa, senza la necessità di essere tutte presenti nel dizionario. Dall’altro, ha reso necessario una revisione del dizionario ottenuto dopo la pura registrazione delle frasi attività, per evitare che la presenza di articoli o verbi ausiliari potesse comportare la segnalazione da parte del software, a seguito della ricerca, di una serie di attività che avessero in comune solo questi elementi con quella da codificare: quindi per esempio, nel dizionario non doveva essere presente l’espressione “**sto** guidando l’auto” ma “guidare auto”. La ricerca basata sul confronto tra insiemi di trigrammi ha infine comportato la necessità di inserire tutte le flessioni dei verbi irregolari con diversa radice (come “cuocere” - “cotto”, “andare” - “vado”, ecc.) perché non si perdessero associazioni solo a causa della diversa forma del verbo utilizzato.

Anche la classificazione ha richiesto delle modifiche (vedi esempio 1) prima di essere utilizzata all’interno del sistema; era infatti necessario che le descrizioni fossero il più esplicative possibile e che i codici più generici contenessero chiari riferimenti ai codici più dettagliati ad essi sottostanti.

Esempio 1: Modifica della classificazione EUROSTAT

CLASSIFICAZIONE EUROSTAT

CLASSIFICAZIONE PER BLAISE

01 dormire

01 dormire, stare al letto malato (il “riposo” va codificato con 032)

011 sonno

011 dormire

012 stare a letto malato

012 stare a letto malato

Inoltre, poiché la classificazione prevede oltre ai semplici codici e alla loro descrizione, anche numerose note che facilitano il codificatore nell’interpretazione di attività che hanno significati molteplici a seconda del contesto (vedi esempio 2) è stato necessario prevedere una collocazione di tali informazioni. Se passare ad un codifica assistita on line ne avesse comportato la perdita infatti, lo strumento ottenuto non si sarebbe rivelato del tutto efficace.

Esempio 2: Note e specifiche

3.1.3.1 AIUTARE APPARECCHIARE, SPARECCHIARE (come aiuto a un familiare convivente) 4.2.1 AIUTARE APPARECCHIARE, SPARECCHIARE (come aiuto a persone non conviventi)

Per inserire nel sistema queste importanti informazioni sono state adottate più soluzioni:

- Introdurre un apposito campo “Nota” nel dizionario, contenente delle specifiche sintetiche che spieghino l’uso corretto del codice (vedi fig. 3)
- Segnalare nell’albero dei codici gerarchici altre sezioni relative ad attività analoghe (vedi fig. 2 - segnalazione sul codice 4.1)

In conclusione il sistema di help, composto dal dizionario e dall’albero dei codici, costituisce un bagaglio informativo completo, necessario allo svolgimento di una corretta attività di attribuzione dei codici.

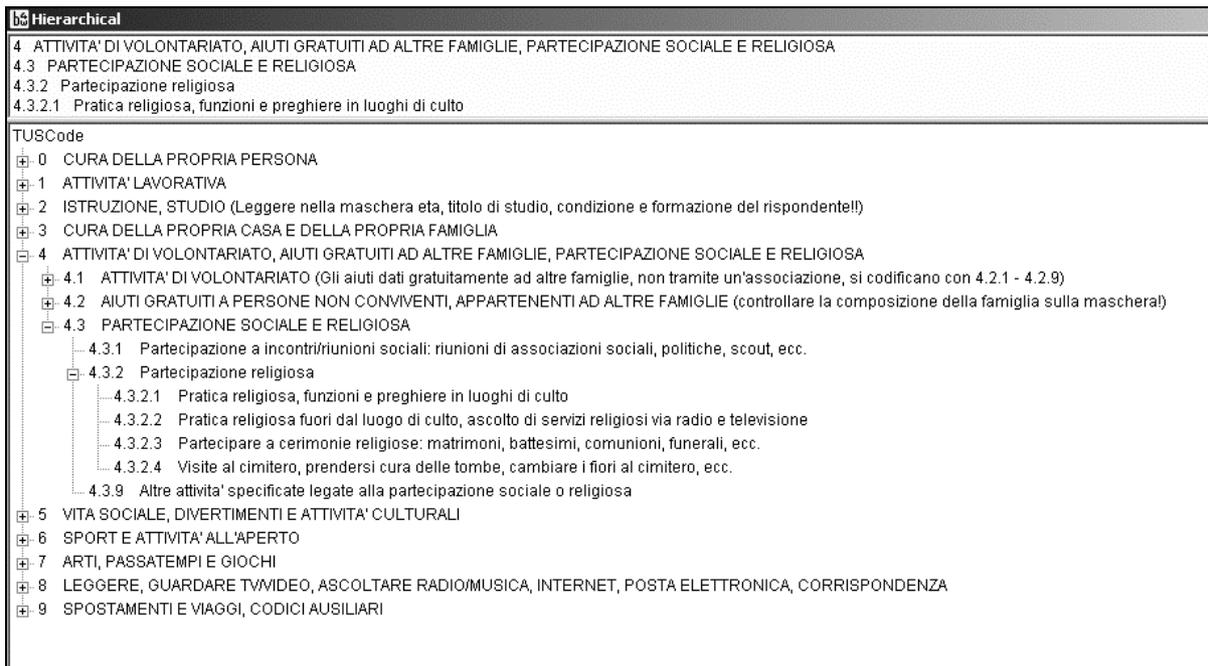


Figura 2. L'albero dei codici

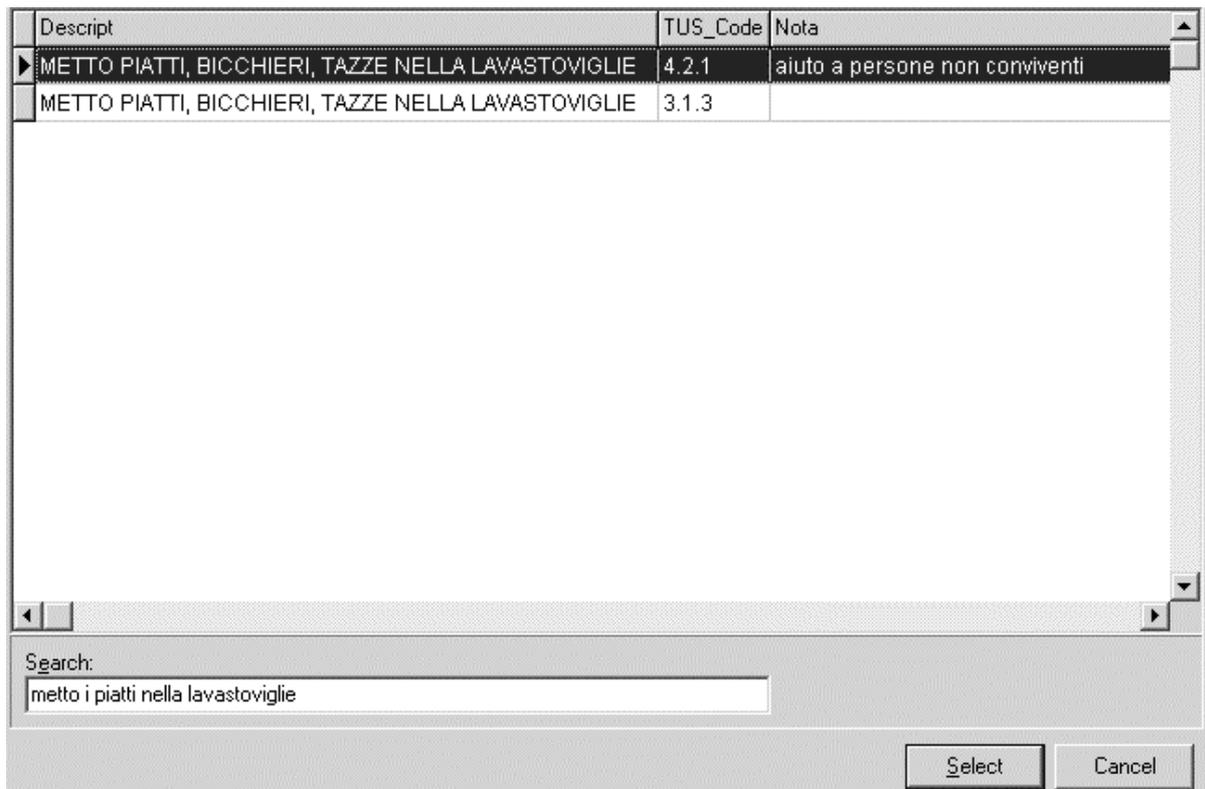


Figura 3. Il dizionario delle attività

6. La maschera Blaise per la codifica

Operativamente dunque, è stata progettata una maschera per la codifica (fig. 4-6) che ha consentito al codificatore di avere a disposizione tutte le informazioni di contesto necessarie allo svolgimento corretto del lavoro, ovvero le caratteristiche socio-anagrafiche sulla persona di cui doveva codificare la giornata, come la professione, l'età, il sesso ecc., alcune informazioni sulle caratteristiche della giornata (data, giorno della settimana), la composizione della famiglia, nonché la sequenza delle attività da registrare.

Il codificatore era tenuto a leggere attentamente le informazioni suddette e a scorrere rapidamente le attività da codificare in modo da formarsi un quadro di insieme della giornata della persona, indispensabile per la corretta traduzione delle attività in codici. Dopo aver letto le informazioni disponibili sulla persona di cui si sta codificando il diario, il codificatore poteva passare alla codifica del primo episodio registrato. Tramite il tasto F1, poteva dare avvio alla ricerca del codice sul sistema di help. La prima ricerca da tentare era quella sul dizionario, che suggerirà il codice in base alla similitudine (a gruppi di tre lettere) della frase registrata con quelle presenti nel vocabolario appositamente costruito. Se la frase era presente ed univocamente associata ad un codice, il codificatore accettava il codice e passava all'attività successiva, se invece il software suggeriva più alternative, il codificatore doveva interpretare il contesto per selezionare il codice corretto.

Nel caso in cui la frase registrata non trovasse una corrispondenza plausibile con nessun'altra presente nel dizionario, il codificatore poteva avviare la ricerca sull'albero dei codici, strutturato secondo la tradizionale forma gerarchica. In questo caso il codificatore poteva navigare l'albero in modo simile alla consultazione dell'albero delle directory in Windows, alla ricerca dell'esatto codice da attribuire all'attività registrata.

Codice codificatore	10
Diario n.	344
Data compilazione	04/02/03
Giorno della settimana	Martedì
Relazione di parentela con 01	Figlio di PR - unico/ultimo matrimonio
Sesso	Maschio
Età	031
Stato civile	Celibe/nubile
Titolo di studio	Diploma media inferiore
Iscrizione corsi scolastici/universitari	No
Condizione	Occupato
Professione	IDRAULICO IN AZIENDA DISTRIBUTRICE ACQUA POTABILE
Settore di lavoro	Commercio all'ingrosso e al dettaglio
N. Componenti della famiglia	3

Figura 4. La maschera per la codifica assistita: le informazioni sull'individuo

	Relazione di parentela con 01	Sesso	Età
(1)	Persona di riferimento	Maschio	054
(2)	Coniuge di PR	Femmina	051
(3)	Figlio di PR - unico/ultimo matrimonio	Maschio	031

PERSONA CHE HA COMPILATO IL DIARIO DA CODIFICARE

Figura 5. La maschera per la codifica assistita: le informazioni sulla famiglia

Indagine Time-Use 2002: Programma per la codifica del diario														
Diari Aiuto														
	Inizio	Fine	Attività principale	Cod. a.p.	Attività contemporanea	Cod. a.c.	Luogo/mezzo spost.	Cod	solo	f<=10	f>=10	fnc	altro	Aus.
[1]	04.00	06.40	HO DORMITO	0.1.1			A CASA	11						
[2]	06.40	07.00	MI SONO ALZATA HO FATTO COLAZIONE	0.2.1.1			A CASA	11			3			11
[3]	07.00	07.10	SONO ANDATO AL LAVORO	9.0.1.3			IN MACCHINA	53			3			
[4]	07.10	07.20	HO LAVORATO	1.1.1			AL LAVORO (CANTI	16	1					
[5]	07.20	13.40	HO LAVORATO	1.1.1			AL LAVORO (CANTI	16						
[6]	13.40	13.50	SONO TORNATO A CASA	9.0.1.3			IN MACCHINA	53			3	4		
[7]	13.50	14.30	HO PRANZATO	0.2.1.1	HO GUARDATO LA TELEVISIONE	8.2.1	A CASA	11			3	4		
[8]	14.30	15.40	MI SONO RIPOSATO	0.3.2	HO GUARDATO LA TELEVISIONE	8.2.1	A CASA	11			3	4		
[9]	15.40	16.20	HO SEGATO LA LEGNA	3.2.3			A CASA	11			3	4		
[10]	16.20	17.00	HO SISTEMATO LA LEGNA	3.2.3	HO PARLATO CON UN AMICO	5.1.2	A CASA	11			3	4		
[11]	17.00	17.30	HO SISTEMATO LA LEGNA	3.2.3			A CASA	11			3	4		
[12]	17.30	18.00	HO PULITO FUORI	3.2.2			A CASA	11			3	4		
[13]	18.00	18.30	MI SONO LAVATO	0.3.1			A CASA	11			3	4		
[14]	18.30	19.20	HO GUARDATO LA TELEVISIONE	8.2.1			A CASA	11			3	4		
[15]	19.20	19.30	HO CENATO	0.2.1.1	HO GUARDATO LA TELEVISIONE	8.2.1	A CASA	11			3	4		
[16]	19.30	20.10	HO CENATO	0.2.1.1			A CASA	11			3	4		
[17]	20.10	21.00	HO GUARDATO LA TELEVISIONE	8.2.1			A CASA	11			3	4		
[18]	21.00	21.10	SONO ANDATO AL BAR	9.0.5.1			IN MACCHINA	53	1					
[19]	21.10	21.40	HO PARLATO CON AMICI	5.1.6			AL BAR	22					5	
[20]	21.40	22.20	HO GIOCATO A CARTE	7.3.2			AL BAR	22					5	
[21]	22.20	23.00	HO PARLATO CON AMICI	5.1.6			AL BAR	22					5	
[22]	23.00	23.10	SONO TORNATO A CASA	9.0.5.1			IN MACCHINA	53	1					
[23]	23.10	04.00	HO DORMITO	0.1.1			A CASA	11						

Figura 6. La maschera per la codifica assistita: il diario

7. Riflessioni conclusive

La disponibilità dell'intero archivio dati di questa indagine oltre che rendere possibili un più ampio e innovativo ventaglio di possibilità di analisi, fornisce una base informativa importantissima ai fini di una riprogettazione metodologica delle rilevazioni sull'Uso del tempo.

Risulta chiaro infatti che a dispetto del formato del diario e delle istruzioni allegate, i rispondenti scelgono liberamente come esprimere quello che fanno e per quanto tempo.

Talvolta appare un vero e proprio gap comunicativo tra ricercatori e rispondenti, perché difficilmente questi ultimi hanno la pazienza di leggere e seguire le istruzioni. Si assiste più che ad una "comunicazione" ad una sorta di "espressione" del vissuto: nel senso che i rispondenti esprimono spesso quello che pensano ci si aspetti da loro, facendone una sintesi personale, seguendo la loro percezione e utilizzando il linguaggio che ritengono più appropriato. Infine, il tempo, non è necessariamente un percorso ripercorribile attraverso uno schema orario, la percezione rende il tempo elastico e le attività non sono sempre svolte l'una dopo l'altra. Gli individui hanno una diversa memoria della durata, una differente percezione della segmentazione del tempo in minuti e ore, una differente propensione a svolgere attività simultanee.

L'archivio testuale potrà consentire di individuare le principali tipologie di questo gap comunicativo, di analizzarlo sul piano quantitativo e fornire strumenti per migliorare l'attuale tecnica di rilevazione.

Per quanto concerne invece l'utilizzo della codifica assistita, i principali vantaggi che ne sono derivati sono sintetizzabili come segue:

- riduzione del numero di codificatori e conseguente maggiore efficacia nel processo di formazione e supervisione degli stessi (sono state previste esercitazioni pratiche e periodici de-briefing);

- monitoraggio quotidiano/settimanale, attraverso gli indicatori appositamente calcolati, dell'operato dei singoli codificatori e della loro rispondenza agli standard qualitativi prefissati;
- supervisione nella codifica di attività di difficile interpretazione;
- individuazione tempestiva di errori sistematici e conseguente adozione dei correttivi più appropriati.

In fase di progettazione si era previsto di aggiornare e incrementare periodicamente il dizionario inizialmente stilato sulla base della pilota, alla luce delle espressioni nuove e frequenti ritrovate nell'indagine in oggetto. Tuttavia, tale aggiornamento non si è reso necessario, data la soddisfacente esaustività del dizionario originario e il carattere altamente ripetitivo delle espressioni usate per descrivere la quotidianità.

Purtroppo il software Blaise non consente di calcolare in automatico le statistiche sulle diverse strategie di codifica adottate dagli operatori, ovvero in quanti casi il processo di codifica è avvenuto tramite il ricorso al dizionario, e in quanti, invece, questo si è rivelato inadeguato ed è stato necessario ricorrere all'albero gerarchico della classificazione.

Tuttavia sperimentazioni da noi effettuate su campioni di diari hanno fatto emergere che la ricerca tramite dizionario dà un esito soddisfacente ed esauriente in circa la metà delle attività da codificare.

Tanto basterebbe a confermare l'impegno necessario ad implementare un processo di codifica semiautomatica: poiché anche qualora essa non dovesse essere di aiuto nella codifica di attività rare o di significato ambiguo, comunque consente di concentrare gli sforzi interpretativi del codificatore solo su di esse.

Bibliografia

- Balboni P.E. (1999). *Parole comuni culture diverse, guida alla comunicazione interculturale*. Saggi Marsilio.
- Camporese R. e Ranaldi R. (2001). Time Use Activities: Translation from Sentences to Codes. Paper presentato alla Riunione scientifica del gruppo SIS (CLADAG), 5-6 luglio, Palermo.
- Camporese R. e Romano M.C. (2003). Children's Time. Paper presentato alla Conferenza IATUR, 17-19 Settembre 2003, Bruxelles.
- Cappadozzi T., Romano M.C., Camporese R. e Vitaletti S. (2001). Assisted coding process: an experiment in progress. Paper presentato alla Conferenza IATUR, ottobre 2001, Oslo.
- Eurostat (1999). *Survey on time use: activity coding list with coding index and coding diary*. DOC E2/TUS/3.2/99.
- Istat (2002a). *Uso del tempo 2002-2003. Istruzioni per la codifica delle informazioni testuali rilevate tramite il diario giornaliero*. Documento ciclostilato.
- Istat (2002b). *Uso del tempo 2002-2003. Istruzioni per l'uso della maschera di codifica del diario giornaliero*. Documento ciclostilato.
- Newmark P. (1988). *La traduzione: problemi e metodi*. Garzanti Ed.
- Romano M.C. (2003). *Le indagini Multiscopo Istat "Uso del tempo"*. In Fraire M. (Ed.), *I Bilanci del Tempo e le indagini sull'Uso del Tempo (Time-Budgets Studies -TBS and Time-Use Surveys-TUS), Metodologie di rilevazione e analisi statistica dei dati sull'uso del tempo umano giornaliero*. Ed. CISU (Centro d'Informazione e Stampa Universitaria).
- Shaw Valerio C. (1993). *Tips for translators*. Bulzoni Ed.