

# **Una soluzione per la rilevazione e codifica della Professione nelle indagini CATI**

## **Autori**

*Stefania Macchia \**, *Manuela Murgia\**, *Loredana Mazza\**, *Giorgia Simeoni\**,  
*Francesca Di Patrizio\*\**, *Valentino Parisi\*\**, *Roberto Petrillo\*\**, *Paola Ungaro\*\**

**2005**

---

\* ISTAT – Servizio Metodologie e Tecniche per la Produzione Statistica  
\*\*ISTAT -- Servizio Popolazione, Istruzione e Cultura

## Indice<sup>1</sup>

<b>1</b>	<b>La rilevazione ed il trattamento della variabile <i>Professione</i></b>	<b>5</b>
<b>2</b>	<b>La codifica della <i>Professione</i> nell'ambito della '<i>strategia in-house</i>' per le rilevazioni CATI</b>	<b>5</b>
<b>3</b>	<b>Il modulo BLAISE per la codifica assistita</b>	<b>7</b>
3.1	<i>L'ambiente di codifica assistita della Professione</i>	8
<b>4</b>	<b>Come rilevare la <i>Professione</i></b>	<b>12</b>
4.1	<i>La logica della Classificazione ufficiale delle Professioni</i>	12
4.2	<i>Perché è difficile rilevare la Professione</i>	13
4.3	<i>Come migliorare la rilevazione della Professione</i>	13
<b>5</b>	<b>L'applicazione di codifica assistita implementata per la codifica della <i>Professione</i></b>	<b>14</b>
<b>6</b>	<b>La formazione degli Operatori CATI</b>	<b>17</b>
6.1	<i>Formazione sulla logica della Classificazione ufficiale</i>	17
6.2	<i>La formazione rivolta all'utilizzo delle stringhe/parole chiave</i>	18
6.3	<i>Formazione rivolta all'utilizzo dei Campi Note</i>	18
6.4	<i>Formazione sugli aspetti tecnici</i>	20
<b>7</b>	<b>Test sull'efficacia della formazione rivolta agli Operatori CATI</b>	<b>21</b>
<b>8</b>	<b>Le sperimentazioni propedeutiche all'adozione del modulo della codifica assistita nelle indagini CATI sulla Transizione istruzione-lavoro</b>	<b>26</b>
8.1	<i>Il sistema di indagini sulla Transizione istruzione-lavoro</i>	26
8.2	<i>Come si è pervenuti all'adozione del modulo di codifica assistita?</i>	27
8.3	<i>Prima fase di sperimentazione: codifica a posteriori con il modulo BLAISE</i>	28
8.4	<i>Seconda fase di sperimentazione: codifica on-line</i>	31
8.5	<i>Conclusioni</i>	34
<b>9</b>	<b>Il monitoraggio della codifica assistita della <i>Professione</i> nelle indagini CATI attraverso le carte di controllo</b>	<b>35</b>
9.1	<i>Le carte di controllo</i>	35
9.2	<i>La procedura di monitoraggio</i>	38
9.2.1	<i>Approfondimento per operatore fuori controllo per '<i>Durata Codifica Professione</i>'</i>	41
9.2.2	<i>Approfondimento per operatore in caso di fuori controllo per '<i>Qualità Codifica Professione</i>'</i>	42
9.2.3	<i>Approfondimento per operatore fuori controllo per '<i>Tasso di Risposta Basso</i>'</i>	42
9.2.4	<i>La procedura di monitoraggio in termini applicativi</i>	45
9.3	<i>La procedura in SAS</i>	47
9.4	<i>Conclusioni</i>	52
<b>10</b>	<b>L'esperienza dell'adozione del modulo di codifica assistita: le Indagini sulla <i>Transizione istruzione-lavoro</i></b>	<b>54</b>

---

<sup>1</sup> Il lavoro è frutto dell'attività di ricerca congiunta degli autori. In ogni caso, ai soli fini dell'attribuzione, i capitoli numero 1, 2, 3 sono da attribuirsi a S. Macchia; i capitoli numero 4, 5, 6, 7 a L. Mazza; i paragrafi numero 8.1 e 8.2 a P. Ungaro; i paragrafi numero 8.4 e 8.5 a F. Di Patrizio; il paragrafo 8.3 a P. Ungaro e F. Di Patrizio; i paragrafi numero 9.1 e 9.3 a G. Simeoni; i paragrafi numero 9.2, 9.2.1, 9.2.2, 9.2.3, 9.2.4, 9.4 a M. Murgia; i paragrafi numero 10.1, 10.2, 10.2.1 a R. Petrillo; i paragrafi numero 10.2.2, 10.2.3, 10.3, 10.4 a V. Parisi.

<i>10.1</i>	<i>I principali risultati delle Indagini 2004 sulla Transizione istruzione-lavoro</i>	<i>54</i>
<i>10.2</i>	<i>Il monitoraggio della codifica della Professione</i>	<i>56</i>
<i>10.2.1</i>	<i>La durata della codifica della Professione</i>	<i>56</i>
<i>10.2.2</i>	<i>La codifica della Professione in corso d'intervista e a posteriori</i>	<i>59</i>
<i>10.2.3</i>	<i>Il dettaglio della codifica delle Professioni rilevate</i>	<i>61</i>
<i>10.3</i>	<i>La qualità della codifica delle Professioni rilevate</i>	<i>62</i>
<i>10.4</i>	<i>Analisi delle Professioni rilevate per Grande Gruppo</i>	<i>63</i>
	<b>Conclusioni</b>	<b>68</b>
	<b>Bibliografia</b>	<b>69</b>

## ***Sommario***

Il lavoro delinea una soluzione per la rilevazione e la codifica della variabile Professione nelle indagini CATI. L'obiettivo da realizzare era quello di ottimizzare la qualità dei dati rilevati inerenti questa variabile, sfruttando l'interazione rilevatore-intervistato ed avvalendosi di un sistema *software* di codifica assistita, tentando nel contempo di limitare al massimo l'appesantimento dell'intervista telefonica. La soluzione descritta è stata individuata a seguito di una serie di sperimentazioni ed è stata adottata nel corso di due Indagini CATI nell'ambito delle quali la *Professione* è una variabile estremamente strategica: *Inserimento professionale dei laureati* e *Percorsi di studio e di lavoro dei diplomati*. Il documento descrive quindi l'intero processo implementato per il trattamento della *Professione* in queste due Indagini, che include alcune applicazioni messe a punto per il monitoraggio del processo stesso, e riporta i principali risultati ottenuti in termini di indicatori della fase di rilevazione CATI.

## ***Abstract***

This job describes a solution to collect and code the Occupation variable in CATI surveys. The aim was to optimise the quality of data concerning this variable, taking advantage of the interaction interviewer-respondent and using a computer assisted software system, trying in the meanwhile to limit the burdening of the telephone interview. This solution has been identified following a series of tests and carried out in two CATI surveys, in which Occupation is an extremely strategic variable: "Upper secondary school graduates" and "University-to-work transition survey and perspectives". The document describes the entire process put in practice for the treatment of Occupation in these surveys, which includes some applications developed for the process monitoring; it also reports the main results obtained in terms of indicators of the CATI interviewing phase.

## 1 La rilevazione ed il trattamento della variabile *Professione*

Come è noto, la variabile *Professione*, rilevata in molte indagini socio-demografiche, è caratterizzata da una notevole complessità sotto diversi aspetti (2003, S. Macchia, L. Mazza, P. Scalisi), quali:

- il *wording* del quesito, tramite il quale, in modo sintetico e chiaro, si deve riuscire a far comprendere al rispondente quali siano gli elementi caratterizzanti che egli deve utilizzare per descrivere la propria *Professione*;
- la formazione degli intervistatori su questa variabile (qualora non sia adottata una tecnica che prevede l'autocompilazione del questionario) che deve essere effettuata in modo che chi effettua l'intervista riesca a farsi fornire dall'intervistato tutti i già citati elementi salienti;
- l'attribuzione alle risposte testuali di un codice secondo una Classificazione predefinita, in modo da rendere elaborabile l'informazione rilevata.

La soluzione individuata in ISTAT ormai da alcuni anni, finalizzata ad ottimizzare l'attività di codifica citata nel terzo punto tramite la minimizzazione degli errori nell'assegnazione dei codici (2000, De Angelis R., Macchia S. and Mazza L.), è quella dell'automazione del processo di codifica ottenuta avvalendosi di opportuni sistemi *software*.

In particolare, in quei casi nei quali l'attività di codifica viene effettuata in una fase successiva rispetto alla rilevazione, è stata messa a punto un'applicazione basata sull'utilizzo del software ACTR (Automatic Coding by Text Recognition); trattasi di un sistema di codifica automatica (l'analisi delle risposte testuali finalizzata all'attribuzione del codice è effettuata in *batch*), già utilizzato in diverse indagini ISTAT, tra le quali il *Censimento della Popolazione del 2001* (2004, S. Macchia, S. Mastroluca).

Il presente lavoro tratta invece della soluzione individuata per le rilevazioni CATI (Computer Assisted Telephone Interviewing), estendibile alle CAPI (Computer Assisted Personal Interviewing), per le quali l'attività di codifica può essere effettuata in corso di intervista; a tal fine, è stata messa a punto un'applicazione di codifica assistita basata sull'utilizzo del sistema BLAISE, di cui si serve per lo sviluppo dei questionari elettronici. Questa strategia consente di sfruttare al massimo l'interazione intervistatore-rispondente, previa, ovviamente, un'accurata formazione dei rilevatori inerente sia la classificazione di riferimento che il modulo di codifica assistita.

Questa soluzione è stata per la prima volta adottata per le due Indagini sulla *Transizione istruzione -lavoro* (*Indagine sull'inserimento professionale dei laureati* e sui *Percorsi di studio e di lavoro dei diplomati*), in merito alle quali saranno descritti: l'applicazione di codifica implementata, la strategia per la formazione dei rilevatori, le sperimentazioni propedeutiche all'adozione del modulo di codifica assistita, il sistema messo a punto per il monitoraggio della qualità della codifica nonché i principali risultati in termini di impatto sulla rilevazione.

## 2 La codifica della *Professione* nell'ambito della '*strategia in-house*' per le rilevazioni CATI

La soluzione appena delineata si inquadra nell'ottica della cosiddetta '*strategia in house per le rilevazioni CATI*' (2003, C. Castagnaro, M. Lo Conte, S. Macchia, M. Murgia). Tale strategia consiste nell'avvalersi di una ditta esterna per tutte le attività e le risorse (umane e *hardware*) inerenti l'effettuazione delle interviste, ma di realizzare *in-house* tutto il *software* necessario per la rilevazione, composto da una serie di moduli, quali:

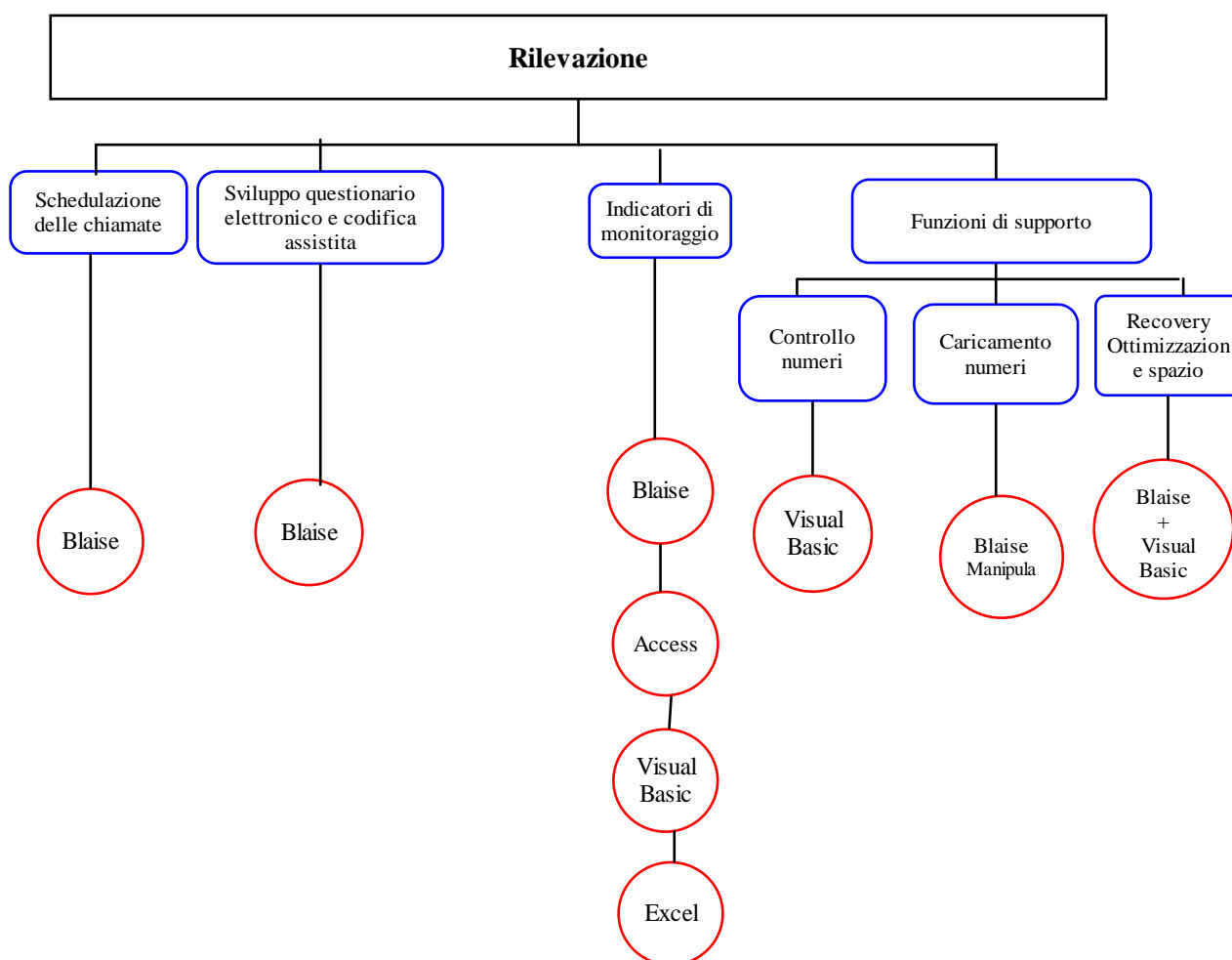
- il questionario di rilevazione;
- il modulo per la schedulazione delle telefonate;
- il sistema di reportistica per il monitoraggio della rilevazione.

Le motivazioni per cui questa strategia si sta diffondendo per diverse indagini dell'Istituto sono molteplici, ma la principale è da ricondursi ad una maggiore garanzia sulla qualità dei dati rilevati, in termini di minimizzazione degli errori non campionari, ottenibile, per l'appunto, tramite l'internalizzazione dello sviluppo del questionario elettronico (corredato da un insieme molto significativo di controlli) e del sistema di indicatori di monitoraggio della rilevazione.

Fino ad oggi questa strategia è stata adottata per l'*Indagine campionaria sulle nascite*, per quella sull'*Inserimento professionale dei laureati* e per quella sui *Percorsi di studio e di lavoro dei diplomati*; è inoltre pianificata per l'*Indagine sulla violenza sulle donne* e per quella sui *Servizi idrici in Italia*.

Dal punto di vista tecnico, come può vedersi dal grafico sottostante, questa strategia integra diversi sistemi *software*, tra i quali il principale è BLAISE.

Figura 2.1 Schema tecnico della 'strategia in-house'



BLAISE è un sistema generalizzato per l'acquisizione dati assistita da computer, sviluppato e commercializzato da Statistics Netherland e già diffuso presso numerosi uffici di statistica non soltanto europei.

E' generalizzato rispetto alla tecnica di rilevazione; è infatti possibile sviluppare questionari elettronici per le seguenti tecniche:

- CATI;
- CAPI;
- CADI (Computer Assisted Data Input);
- CAWI (Computer Assisted Web Interviewing).

BLAISE è composto da un insieme di moduli, tra i quali uno per la codifica assistita della variabili rilevate a testo libero.

### 3 Il modulo BLAISE per la codifica assistita

Il modulo di codifica assistita di BLAISE è anch'esso generalizzato, nel senso che è indipendente dalla lingua e dalla Classificazione di riferimento. In sintesi, consta di un motore per il trattamento dei testi che effettua il *matching* tra la risposta testuale ed i testi della base informativa di riferimento; questo confronto è finalizzato ad individuare il testo della base informativa più vicino a quello da codificare, cui corrisponde quindi il codice da attribuire.

Il fatto che si tratti di un modulo generalizzato implica che, come si vedrà di seguito, la costruzione della citata base informativa propria di ciascuna Classificazione sia a carico dell'utente. Questo patrimonio informativo è costituito da due elementi:

- il cosiddetto 'dizionario' → insieme di descrizioni testuali alle quali sono associati i codici della Classificazione;
- la struttura gerarchia della Classificazione → determinata in funzione della gerarchia tra i codici.

Il dizionario costituisce l'elemento più delicato dell'applicazione, in quanto la sua ricchezza, in termini di numerosità di descrizioni contenute e di terminologia in esse utilizzata, che deve essere quanto più possibile vicina al linguaggio utilizzato dai rispondenti, impatta direttamente sul successo dell'attività di codifica, ossia sulla possibilità di individuare il codice corretto.

Il motore per il trattamento dei testi utilizza una serie di algoritmi basati sulla scomposizione delle stringhe in 'trigrammi'; più in dettaglio, questa scomposizione fa sì che ciascuna parola componente la descrizione produca un'array di trigrammi ed, in particolare, una parola di  $n$  caratteri corrisponda ad una 'working string' di  $n+2$  caratteri e produca un'array di  $n$  trigrammi (si veda figura 3.1).

Figura 3.1 Scomposizione di una stringa in trigrammi

		Working string						
		$W_1$	$W_2$	$W_3$	$W_4$	$W_5$	$W_6$	$W_7$
Array di trigrammi	$T_1$	∅	S	E				
	$T_2$	→	S	E	G			
	$T_3$		→	E	G	N		
	$T_4$			→	G	N	O	
	$T_5$				→	N	O	∅

La figura 3.1 mostra la scomposizione della parola 'SEGNO'. Come si vede vengono prodotti una 'working string' di 7 caratteri ( $W_1 - W_7$ ) e un vettore di 5 trigrammi ( $T_1 - T_5$ ).

Il *matching* testuale avviene in tre fasi:

1) *Search process*: selezione dei trigrammi della descrizione da codificare da ritenersi 'significativi', ai fini del *match*, ed individuazione delle descrizioni del dizionario che hanno in comune con il testo da codificare uno o più di questi trigrammi.

- 2) *Select process*: selezione dei record da estrarre dal dizionario dei quali effettuare il *display*.
- 3) *Display process*: definizione della sequenza con cui deve essere effettuato il *display* dei record selezionati.

I processi decisionali espletati in queste tre fasi vengono effettuati sulla base dei risultati di una serie di algoritmi (1999, BLAISE Developer's Guide) che tengono conto dei seguenti elementi:

- numero di trigrammi della descrizione da codificare;
- numero di record del dizionario;
- per ciascun trigramma, numero dei record del dizionario che lo contengono;
- numero di trigrammi della descrizione da codificare (selezionati nella fase 1) contenuti in ciascun record del dizionario;
- lunghezza (numero di trigrammi) di ciascuna descrizione del dizionario che contiene uno o più trigrammi della descrizione da codificare;
- massimo numero di trigrammi della descrizione da codificare trovati in un record del dizionario.

In sintesi, non ritenendo essere questo il contesto per entrare in dettaglio sul funzionamento del modulo di codifica assistita, si esplicita che il risultato finale del *matching* è il *display* a video di un insieme di record del dizionario in cui ai primi posti figurano quelli con le descrizioni più brevi e con il maggior numero di trigrammi abbinati (in termini tecnici la schermata che appare si chiama di *Look-up*).

La navigazione nel dizionario da parte del codificatore può avvenire secondo tre modalità:

- secondo la dizione alfabetica → il codificatore digita un testo ed il sistema ricerca ed estrae dal dizionario i testi che hanno in comune con quello digitato uno o più trigrammi;
- secondo l'albero → il codificatore visualizza in un primo tempo la Classificazione a livello dei rami gerarchicamente più elevati, quindi seleziona il ramo ed entra nei codici specifici del livello inferiore e così via;
- secondo una procedura mista → il codificatore seleziona un ramo della Classificazione e quindi effettua una ricerca testuale nell'ambito di quel ramo.

L'abilitare la consultazione tramite l'una o l'altra di queste modalità viene deciso in funzione di diversi fattori, quali:

- la struttura della Classificazione (se non ha una struttura gerarchica, la navigazione ad albero non è applicabile);
- la numerosità dei rami dell'albero (se troppo elevata può essere più conveniente non procedere secondo la gerarchia);
- la numerosità delle descrizioni associate a ciascun codice del dizionario (se molto elevata può essere conveniente una navigazione mista che individui il ramo e solo in questa fase effettui una ricerca testuale);
- la facilità da parte del codificatore di individuare correttamente i rami nei quali entrare (un errore sulla scelta del ramo si ripercuote sull'esito della codifica).

### 3.1 L'ambiente di codifica assistita della Professione

Come già accennato, la costruzione dell'ambiente di codifica, soprattutto per quanto attiene la predisposizione ed arricchimento del dizionario è una fase molto delicata completamente a carico dell'utente.

Un dizionario per la codifica, sia essa automatica o assistita, per poter essere elaborato dal *software*, deve contenere esclusivamente descrizioni sintetiche, analitiche e non ambigue.



Per raggiungere questo obiettivo, la prima attività da realizzare consiste nella rielaborazione del manuale ufficiale della Classificazione e la sua integrazione con materiale di riferimento (per esempio documentazioni ufficiali inerenti la Classificazione in esame).

La seconda attività, non meno importante della prima, è invece inerente l'arricchimento del dizionario stesso con risposte empiriche (pre-codificate) fornite dagli intervistati nel corso di precedenti rilevazioni.

I processi da realizzare per la prima tipologia di attività riguardano (2001, Macchia S.):

- la semplificazione delle descrizioni complesse → ossia delle descrizioni che riassumono più concetti associati allo stesso codice;
- la definizione di sinonimi → concetti elementari/specifici che si riferiscono a parole generiche;
- la rielaborazione delle classi aperte;
- l'eliminazione delle clausole di esclusione;
- l'integrazione con materiale di riferimento.

Relativamente alla *Professione*, la base informativa da cui si è partiti è stato il manuale ufficiale della Classificazione ISTAT CP2001. Ciò che può dare un'indicazione circa l'impatto sul dizionario elaborabile della fase di rielaborazione del manuale ufficiale è il fatto che questa abbia comportato un incremento di circa 1000 testi rispetto a quelli originariamente contenuti nel manuale stesso.

Per la seconda attività, inerente l'arricchimento del dizionario con risposte empiriche pre-codificate, sono state utilizzate le seguenti fonti:

- *Indagine di qualità del Censimento della popolazione del 1991*
- *Indagine sulla salute (1994)*
- *Indagine sulle Forze di Lavoro (4 trimestri 1998)*
- *I e II indagine pilota del Censimento della popolazione del 2001*
- *Censimento della Popolazione 2001 (convivenze).*

Operativamente tutte queste attività sono state effettuate in occasione della predisposizione dell'applicazione di codifica automatica in ambiente ACTR, utilizzata in diverse Indagini dell'Istituto, tra le quali il *Censimento della Popolazione*; tuttavia, come già specificato, i processi finora descritti sono comuni alle diverse strategie di codifica.

Le operazioni che sono state necessarie per la migrazione di questa applicazione in ambiente BLAISE sono state quelle derivanti da due fattori:

- 1) il motore per il trattamento dei testi, diverso tra i due *software*;
- 2) il fatto che le due strategie di codifica (automatica/assistita) lavorino con due approcci molto diversi tra loro.

Relativamente al primo dei due fattori, il motore per il trattamento dei testi di ACTR non lavora per trigrammi, ma per stringhe, nell'ambito delle quali individua le parole (1988, M.J. Wenzowski); il *matching* tra le stringhe viene effettuato a seguito di un confronto tra le parole, successivo, rispetto ad alcuni processi di trattamento delle stesse, quali, per esempio, l'eliminazione di parole ininfluenti (articoli, preposizioni, congiunzioni) o l'eliminazione delle componenti che determinano genere e numero. Inoltre, ACTR gestisce i sinonimi (a livello di parola o coppie di parole) come file esterni al dizionario che, in fase di analisi del testo, ACTR sostituisce nell'ambito delle stringhe del dizionario stesso.

Le attività necessarie nel corso della migrazione dell'ambiente di codifica da ACTR a BLAISE derivanti da questi fattori sono state, per esempio:

- l'inserimento, nel dizionario BLAISE, di descrizioni che contenessero i sinonimi gestiti da ACTR con i file

esterni;

- l'eliminazione, nel dizionario BLAISE, di articoli, preposizioni e congiunzioni; questa operazione non è strettamente necessaria, ma è stata opportuna in considerazione della elevata numerosità di descrizioni del dizionario delle *Professioni*; la presenza di questi elementi nella stringa, infatti, avrebbe comportato il *matching* di trigrammi non significativi e quindi la visualizzazione al codificatore di una serie più numerosa di stringhe tra le quali individuare quella cui corrisponde il codice corretto.

In merito ai diversi approcci dei due sistemi di codifica, occorre osservare che un sistema di codifica automatica, quale è ACTR, ha il fine di individuare il singolo testo del dizionario corrispondente al testo da codificare, mentre un sistema di codifica assistita, come BLAISE, ha il fine di selezionare dal dizionario una serie di testi simili a quello da codificare, lasciando poi al codificatore la selezione di quello più simile a quello da codificare.

Questo significa che nel dizionario di ACTR erano state inserite:

- alcune dizioni simili tra di loro per aumentare la possibilità di individuare il *match*;
- alcune dizioni corrispondenti a codici fittizi oppure alla sigla 'nc' (non codificabile), per evitare match errati e perché potessero essere facilmente individuate per essere trattate nella fase di controllo e correzione. In particolare, i codici fittizi erano stati associati a descrizioni generiche, ma molto frequenti, che non consentivano di individuare univocamente il Grande Gruppo (per esempio 'operaio' non rappresenta una *Professione* ed è riconducibile ai Grandi Gruppi 6, 7 e 8); la sigla 'n.c.' era stata associata a descrizioni che non rappresentavano una situazione lavorativa, ma che erano state fornite dal rispondente per un errore di percorso del questionario (per esempio: 'pensionato' oppure 'casalinga').

In fase di migrazione in ambiente BLAISE, le descrizioni associate alla sigla 'n.c.' sono state eliminate, perché, data la tecnica CATI, non erano possibili errori di percorso e, comunque, l'intervistatore stesso non avrebbe considerato tali espressioni essere risposte valide per descrivere la *Professione*; le descrizioni corrispondenti a codici fittizi, invece, sono state ripetute in corrispondenza di ciascun Grande Gruppo, associando a ciascuna di esse una nota esplicativa.

Infine, proprio per guidare il codificatore nel rilevare le informazioni necessarie per individuare la *Professione* al massimo dettaglio, sono state inserite apposite 'note' in corrispondenza delle dizioni associate a codici non completi; in pratica, le note riportate danno un'indicazione sugli elementi da richiedere all'intervistato per proseguire nell'individuazione della *Professione* (cfr. capitolo 6).

A termine di queste attività, la dimensione del dizionario della *Professione* in ambiente BLAISE in termini di numero di descrizioni è quella riportata nella seguente tabella.

Tabella 3.2 Dimensione del dizionario conseguente alle diverse fasi di arricchimento

Numero di descrizioni del manuale ufficiale della Classificazione (Voci professionali)	Numero di descrizioni del dizionario elaborabile
6.300	16.125

Infine, al momento dell'adozione di questo modulo per le due Indagini sulla *Transizione istruzione-lavoro*, è stato necessario scegliere quale modalità di navigazione nel dizionario dovesse essere resa disponibile e predisporre in tal senso l'applicazione.

A seguito di una sperimentazione effettuata prima dello svolgimento dell'Indagine, è stato ritenuto preferibile

predisporre la navigazione secondo la dizione alfabetica; la motivazione principale è da ricondursi al fatto che, vista la complessità dei criteri classificatori per l'individuazione dei Grandi Gruppi, già la selezione del ramo gerarchicamente superiore avrebbe presentato una serie di problematiche che avrebbero rischiato di compromettere l'individuazione del codice corretto.

Inoltre, ritenendo che il tempo necessario per l'individuazione del codice fosse un fattore cruciale, in quanto un tempo eccessivo avrebbe potuto mettere a repentaglio la conclusione dell'intervista, è stato pensato di dare all'intervistatore una duplice opportunità: codificare nel corso dell'intervista, oppure, qualora ci si accorgesse di aver speso più di due minuti su tale attività senza riuscire nell'individuazione del codice completo a quattro digit, di rimandare il lavoro a fine intervista.

Per garantire tali possibilità, sono stati implementati nel questionario elettronico due accorgimenti:

- al passare dei due minuti, è stato fatto apparire a video un messaggio per l'intervistatore che lo invitava a proseguire con l'intervista;
- a fine intervista è stato riproposto il modulo di codifica assistita, corredando la schermata con una serie di informazioni, rilevate durante l'intervista stessa, che potevano essere di supporto per la codifica della *Professione* (Es. posizione della professione, settore dell'attività economica, ecc.).

Infine, la funzione di codifica a fine intervista è stata automaticamente attivata ogni qual volta l'intervistatore non riuscisse ad individuare, nel corso dell'intervista, un codice completo (a 4 digit).

Nel capitolo 8 viene descritta in dettaglio la sperimentazione propedeutica alla scelta della modalità di navigazione del dizionario.

## 4 Come rilevare la *Professione*

### 4.1 La logica della *Classificazione ufficiale delle Professioni*

La *Professione* viene codificata in ISTAT tramite la *Classificazione ufficiale* (Metodi e Norme n. 12 2001). La logica della *Classificazione ufficiale* riprende sostanzialmente quella internazionale della ISCO 88, basata sul criterio della *competenza (skill)*, intesa nella sua duplice dimensione: il *livello* delle competenze (*skill level*) e il *campo* di applicazione delle competenze (*skill specialization*). La prima dimensione coglie una differenza gerarchica tra le professioni ed è assimilabile, sostanzialmente, al livello di istruzione formale necessario allo svolgimento di una data professione. La seconda, invece, consente una articolazione orizzontale delle professioni e viene usata principalmente per cogliere le differenze interne ai Grandi Gruppi in relazione alle differenze settoriali, alle attrezzature utilizzate, ai materiali lavorati, alla natura dei servizi prodotti e ad altre caratteristiche specifiche dell'ambito in cui si svolgono le diverse professioni. Altri criteri sottesi dalla *Classificazione* sono:

- il livello di *responsabilità*;
- il grado di *autonomia*;
- la *complessità* del lavoro;
- la componente *manuale/intellettuale* delle mansioni.

La *Professione* viene così codificata sulla base di una *Classificazione ufficiale* strutturata in 9 Grandi Gruppi definiti in base a quattro livelli di istruzione (4 laurea, 3 diploma di scuola secondaria superiore, 2 obbligo scolastico, 1 alfabetizzazione di base) come mostra la tabella che segue:

Tabella 4.1 – I nove Grandi Gruppi della *Classificazione ufficiale*

<i>Grandi Gruppi (*)</i>	<i>Livello di istruzione</i>
1 Legislatori dirigenti imprenditori	-
2 Professioni intellettuali, scientifiche e di elevata specializzazione	4
3 Professioni tecniche	3
4 Impiegati	2
5 Professioni qualificate nelle attività commerciali e nei servizi	2
6 Artigiani, operai specializzati e agricoltori	2
7 Conduttori di impianti e operai semiqualeficati addetti a macchinari fissi e mobili	2
8 Professioni non qualificate	1
9 Forze armate	-

(\*) Il primo e il nono Grande Gruppo comprendono delle Professioni collocate al di fuori della struttura gerarchica della *Classificazione*, in quanto contemplano dei livelli di istruzione particolari.

Procedendo con un dettaglio maggiore, ogni Grande Gruppo è articolato in Gruppi, Classe nel Gruppo e Categoria nella Classe. La codifica dei Grandi Gruppi, dei Gruppi, delle Classi e delle Categorie professionali segue una

numerazione decimale articolata su quattro codici (digit). Nel suo complesso la Classificazione ufficiale comprende 519 Categorie alle quali deve essere ricondotto l'universo non quantificabile delle *Professioni*.

#### 4.2 Perché è difficile rilevare la Professione

Descrivere ed identificare correttamente la propria *Professione* spesso risulta non facile né tanto meno immediato. Da analisi condotte su risposte rilevate a testo libero, in precedenti Indagini, sul quesito della *Professione* si è visto che non sempre il rispondente sa esplicitare in maniera esauriente in cosa consiste la propria attività lavorativa. Dall'analisi dei testi, infatti, si è verificato che chi risponde:

- descrive solo la **posizione che occupa nella professione**, ad esempio *dirigente, operaio, impiegato*;
- spesso espleta **mansioni afferenti a professioni diverse**, ad esempio *impiegato addetto alle vendite, impiegato contabile amministrativo*;
- non specifica il **livello di responsabilità**, ad esempio *impiegato ufficio del personale, impiegato segreteria*;
- non dichiara la **specializzazione della competenza**, nonché il livello di istruzione e la preparazione tecnica raggiunta ad esempio *operaio edile, operaio agricolo*.

Vengono cioè a mancare tutte o in parte quelle informazioni che risultano indispensabili per una codifica al massimo dettaglio, considerati i criteri strutturali propri della Classificazione ufficiale. Il rischio di incorrere in una codifica non corretta oppure di fare forzature, mancando le informazioni necessarie, è pertanto molto elevato.

#### 4.3 Come migliorare la rilevazione della Professione

L'acquisizione corretta delle informazioni di dettaglio è il punto cardine per migliorare la rilevazione della *Professione* e quindi la sua codifica. Relativamente a questo aspetto è saliente il *wording* del quesito che rileva la variabile. Una cura particolare dovrà essere rivolta agli esempi riportati nel questionario che non debbono contenere solo lemmi sintetici come per esempio *falegname, parrucchiere o cardiologo*, facilmente codificabili e univoci all'interno della Classificazione ufficiale, bensì esempi di descrizioni che contengono o fanno intendere la necessità di specificare il grado di specializzazione, la responsabilità e la competenza raggiunta nel campo professionale, ad esempio *'impiegato esecutivo ufficio del personale'* oppure *'impiegato responsabile ufficio del personale'* e non semplicemente *'impiegato ufficio personale'*.

Un altro aspetto saliente per migliorare la rilevazione del fenomeno è sicuramente la formazione dei rilevatori sui criteri strutturali della Classificazione. Ciò consentirà, infatti, all'operatore sia di essere in grado di cogliere e riportare solo gli aspetti significativi delle risposte date dall'intervistato, che eventualmente, di chiedere ulteriori informazioni qualora ritenga che le stesse non siano esaustive per poter procedere alla codifica corretta della *Professione*. Nelle Indagini sulla *Transizione istruzione-lavoro* si è lavorato seguendo questa duplice ottica ovvero:

- sono stati accuratamente predisposti e inseriti esempi mirati di *Professioni* nei quesiti che rilevano la variabile;
- gli operatori CATI sono stati formati sia sulla logica della Classificazione ufficiale, che sugli aspetti più propriamente tecnici per l'utilizzo del modulo di codifica assistita.

Nell'*Indagine sui percorsi di studio e di lavoro dei diplomati 2001* il quesito sulla *Professione* recitava, infatti, così:  
***Puoi dirmi cortesemente qual è la tua professione? Ti raccomando di non usare termini generici come impiegato o operaio***

Per l'intervistatore: assicurarsi che l'intervistato indichi sia il livello delle competenze richieste (ad es. tecnico meccanico, operaio meccanico) sia il campo delle competenze (ad es. perito agrario, tecnico elettronico). Specificare il livello di responsabilità associato allo svolgimento della mansione (ad es. impiegato responsabile, impiegato esecutivo). Inoltre, in diversi casi è necessario che l'intervistato indichi il luogo dove svolge la sua attività (ad es. cassiere in banca, cassiere di negozio, professore di scuola media, professore di scuola secondaria superiore).

Come si vede, la nota per l'intervistatore, riportando esempi di *Professioni* dettagliate, ricorda all'operatore come devono essere raccolte le informazioni per arrivare ad una codifica corretta ed univoca. L'adozione poi di un modulo di codifica assistita per l'automatizzazione del processo di attribuzione del codice alla risposta testuale apporta sicuramente un miglioramento in termini di qualità in quanto rende il processo ripetibile e limita la discrezionalità interpretativa del codificatore.

In questa ottica si è collocato l'utilizzo del modulo di codifica assistita in BLAISE all'interno delle due Indagini CATI.

## **5 L'applicazione di codifica assistita implementata per la codifica della *Professione***

Come già è stato detto nei capitoli precedenti, il *software* BLAISE, dispone di un apposito modulo per la codifica assistita che consente la navigazione nel dizionario informatizzato in tre modi differenti:

- *secondo la dizione alfabetica;*
- *secondo l'albero della Classificazione;*
- *secondo una procedura mista.*

A prescindere dalla modalità di ricerca che si decide di adottare, la base informativa è costituita dal dizionario informatizzato. Esso corrisponde fisicamente a un file a struttura sequenziale contenente:

<i>Codice</i>	<i>Descrizione testuale associata</i>	<i>eventuale Campo Note</i>
---------------	---------------------------------------	-----------------------------

Nel caso specifico (*cf. paragrafo 3.1*) si è partiti dal manuale della Classificazione ufficiale 2001. Su di esso è stato necessario effettuare un trattamento dei testi in maniera tale da renderli leggibili dal *software* di codifica. Il trattamento dei testi ha comportato un lavoro che ha consentito di prevedere descrizioni sintetiche, esplicite, non ambigue e vicine al linguaggio utilizzato dai rispondenti.

Il dizionario informatizzato, predisposto per la codifica assistita, contiene le 6300 Voci Professionali proprie della Classificazione ufficiale e circa 10.000 risposte empiriche ricavate da Indagini che rilevano il fenomeno. Lo scopo dell'inserimento dei testi empirici è stato quello di cercare di facilitare e, contemporaneamente, aumentare la possibilità di successo del *match* di codifica, prevedendo lemmi che si avvicinano di più al linguaggio utilizzato dai rispondenti rispetto ai testi presenti nel manuale ufficiale.

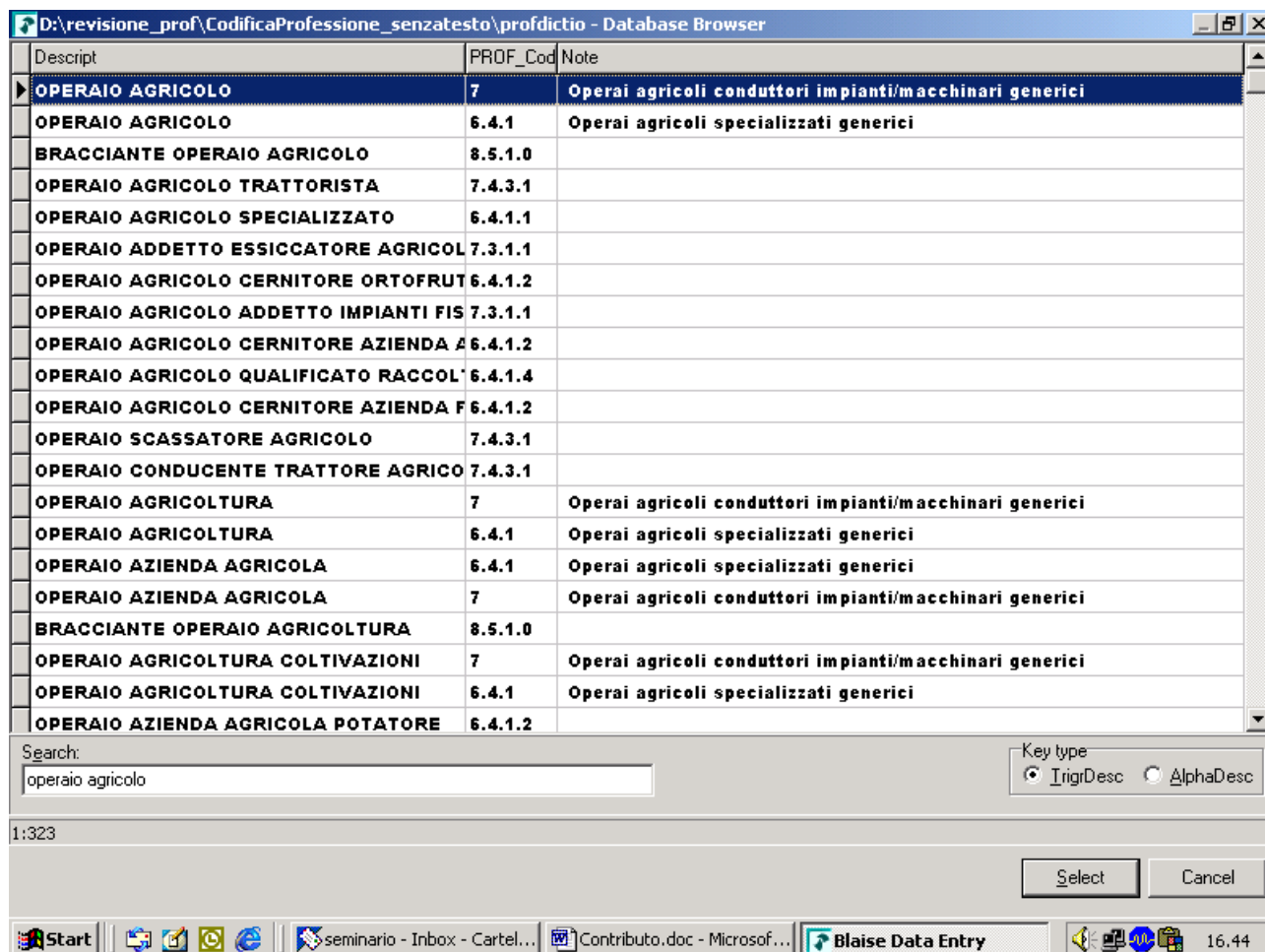
All'interno del dizionario sono presenti sia lemmi corrispondenti a codici a quattro digit (massimo dettaglio) che testi associati a codici con un numero inferiore di digit (codici generici), in quanto le informazioni contenute in essi non consentono di arrivare al massimo dettaglio. Nella costruzione del modulo di codifica assistita in BLAISE e quindi del dizionario informatizzato, si è cercato di ridurre al minimo la possibilità di incorrere in una codifica non

corretta, da parte del codificatore, oppure di fare forzature nell'attribuzione del codice in caso di risposte ambigue. La logica su cui è stata fondata tutta la costruzione dell'applicazione di codifica assistita è stata, pertanto, quella di cercare di essere d'aiusilio all'operatore per aiutarlo nel suo delicato lavoro, soprattutto per dirimere l'ambiguità della codifica di *Professioni* generiche non codificabili univocamente. Per questo scopo, a fianco dei testi con codici minori di quattro digit sono stati inseriti dei **'Campi Note'**. Le funzioni dei Campi Note diventano allora quelle di:

- avvisare l'operatore che si sta rilevando una *Professione* generica;
- segnalare la necessità di chiedere all'intervistato ulteriori specifiche sulla *Professione* dichiarata per arrivare ad una codifica al massimo dettaglio, in corso d'intervista;
- porre una maggiore attenzione nell'attribuzione dei digit mancanti, qualora l'operatore stia effettuando una codifica assistita post intervista.

Segue pertanto che i testi relativi a *Professioni* generiche, passibili cioè di essere codificate anche all'interno di Grandi Gruppi diversi, siano presenti nel dizionario informatizzato associati ai codici diversi corrispondenti a ciascun Grande Gruppo. Questo per consentire all'operatore, trovandosi davanti la casistica completa delle varie possibilità di codifica, di poter effettuare la scelta corretta dopo aver individuato le informazioni mancanti.

Per chiarire meglio, consideriamo per esempio la *Professione* generica 'operaio agricolo' la schermata che appare all'operatore è la seguente:



Come si vede, formulando domande mirate all'intervistato in corso di intervista, avvalendosi anche dell'aiusilio del *Campo Note*, e riportando le informazioni di dettaglio mancanti nel campo di ricerca *Search* l'operatore potrà

effettuare la ricerca all'interno degli 'Operai agricoli specializzati', tra i 'Conduttori di impianti e macchinari agricoli' oppure codificare come 'Bracciante operaio agricolo' (testo privo del *Campo Note* perché codice completo a 4 digit).

Oltre all'inserimento delle *Note*, un altro accorgimento per minimizzare gli errori di codifica è stato quello relativo all'introduzione, nei testi delle empiriche inserite nel dizionario informatizzato, delle cosiddette '*Stringhe/Parole chiave*'. L'utilizzo delle *parole chiave* nella fase di rilevazione della *Professione*, come verrà specificato meglio nel capitolo successivo, consente all'operatore di discriminare più facilmente la ricerca all'interno dei nove Grandi Gruppi.

Per capire come si è arrivati a determinare ed individuare le *stringhe/parole chiave* possiamo considerare come esempio la codifica delle '*Professioni impiegate*', che la Classificazione ufficiale pone all'interno sia del Grande Gruppo 3 (*Professioni tecniche*) che nel Grande Gruppo 4 (*Impiegati*). In base a criteri classificatori validati dai responsabili della Classificazione ufficiale sono stati considerati sinonimi di:

- *professioni tecniche* (Grande Gruppo 3) → *gli impiegati direttivi, gli impiegati con funzioni di responsabilità o di coordinamento*;
- *impiegati* (Grande Gruppo 4) → *gli impiegati esecutivi, gli impiegati con mansioni generiche o di concetto*;

In questo caso *stringhe/parole chiave* diventano:

- *impiegato tecnico* → *impiegato direttivo* → *impiegato responsabile* → *impiegato coordinatore*;
- *impiegato* → *impiegato esecutivo* → *impiegato di concetto* → *impiegato mansioni generiche*.

Da ultimo, al fine di evitare di appesantire il dizionario con trigrammi inutili ai fini del *match* di codifica e contemporaneamente di alleggerire l'operatore nella scrittura del testo durante l'intervista, nella fase di implementazione del contesto di codifica assistita è stato necessario effettuare degli interventi mirati. Tenendo infatti conto della logica di ricerca utilizzata da BLAISE, si è proceduto ad inserire una *standardizzazione dei testi linguistici* all'interno del dizionario informatizzato di *Look-up*.

Si ricorda infatti, come già è stato detto nei capitoli precedenti, che il motore di ricerca utilizzato da BLAISE per il *match*, è basato sul trattamento dei testi per trigrammi e che gli algoritmi utilizzati nella ricerca tengono conto: del numero di trigrammi della descrizione da codificare, del numero di record del dizionario, per ciascun trigramma del numero dei record del dizionario che lo contengono, della lunghezza del testo....

Gli interventi che sono seguiti hanno, pertanto, determinato:

- l'eliminazione all'interno dei testi di tutte le *parti invariabili del discorso* (preposizioni semplici, articolate, congiunzioni e avverbi);
- l'inserimento delle *Professioni* al genere maschile singolare;
- il non inserimento delle *parti variabili del discorso* come per esempio i verbi e gli articoli;
- l'eliminazione dai testi della punteggiatura, dei segni ortografici (parentesi, virgolette, trattino, \*, \$, /, apostrofo);
- il non utilizzo di abbreviazioni nei testi;
- la correttezza ortografica all'interno dei testi.



## 6 La formazione degli Operatori CATI

La rilevazione e quindi la codifica della variabile *Professione* all'interno delle indagini sulla *Transizione istruzione-lavoro*, rivestono come è noto, una importanza notevole. S'intuisce, pertanto, quanto sia stato delicato e importante il ruolo ricoperto dall'operatore CATI. Egli, in corso d'intervista telefonica, doveva infatti:

- raccogliere dall'intervistato le informazioni di dettaglio relativamente al livello di responsabilità e al tipo di mansioni svolte nell'ambito della *Professione* dichiarata;
- guidare l'intervistato laddove riteneva che le informazioni raccolte fossero carenti per una codifica al massimo dettaglio, minimizzando così l'attribuzione di codici generici.

Si ricorda, infatti, che in entrambe le Indagini la codifica della *Professione* dei ragazzi diplomati/laureati 2001 era richiesta al massimo dettaglio ovvero a quattro digit. Si è ritenuto quindi necessario effettuare una accurata formazione agli operatori sia sulle Indagini in generale che sulla rilevazione/codifica della *Professione* in particolare.

La formazione effettuata sulla rilevazione/codifica della *Professione* è stata rivolta a:

- fornire le conoscenze di base sulla *logica della Classificazione ufficiale* e sui criteri classificatori;
- promuovere l'uso delle *parole chiave* per migliorare e discriminare la ricerca all'interno dei nove Grandi Gruppi;
- insegnare a formulare domande mirate qualora l'intervistato non fornisca subito le informazioni necessarie per arrivare ad una codifica completa (avvalendosi anche dell'ausilio dei *Campi Note*);
- conoscere gli *aspetti più propriamente tecnici* per l'utilizzo del modulo di codifica assistita.

### 6.1 Formazione sulla logica della Classificazione ufficiale

Prima di entrare in merito all'utilizzo del modulo di codifica assistita, è stato necessario illustrare agli operatori sia la logica a cui risponde la Classificazione ufficiale che la struttura gerarchica della stessa. La logica della Classificazione ufficiale (*cfr capitolo 4*), è basata sul criterio della competenza inteso nella sua duplice dimensione del *livello delle competenze* e del *campo di applicazione delle competenze* ed è strutturata in 9 Grandi Gruppi.

Nella fase di formazione rivolta agli operatori per ogni Grande Gruppo:

- sono stati forniti esempi di professioni dettagliate (ovvero con le informazioni di dettaglio complete), avendo cura di non riportare esclusivamente i testi presenti nel manuale ufficiale. Ad esempio nel caso del Grande Gruppo I (*Legislatori, Dirigenti, Imprenditori*) sono stati forniti esempi di *Professioni* che permettono di arrivare ad una codifica univoca a 4 digit come '*deputato*', '*giudice*', '*dirigente grande azienda privata dell'edilizia*', '*piccolo imprenditore di negozio*';
- nel caso di risposte relative a professioni generiche (ovvero con le informazioni di dettaglio non complete), scelte tra le più frequenti all'interno dei vari Grandi Gruppi, sono stati forniti degli esempi di informazioni dettagliate da raccogliere dall'intervistato, qualora non siano state esplicitate dallo stesso. Ad esempio una *Professione* generica molto frequente che ricade all'interno del Grande Gruppo I è quella di 'dirigente'. Le informazioni dettagliate che l'operatore doveva imparare a chiedere qualora non fossero state date erano: '*è un dirigente pubblico o privato?*' qualora fosse un dirigente pubblico è di '*Governo/Amministrazione statale/Magistratura/Organizzazione di interesse nazionale?*'; se invece fosse un dirigente privato se di una *grande o piccola impresa* e infine il settore dove opera l'azienda.

## 6.2 La formazione rivolta all'utilizzo delle stringhe/parole chiave

Come anticipato, i testi delle empiriche inserite nel dizionario informatizzato contengono delle *stringhe/parole chiave* che, se usate e riportate nel testo della *Professione* rilevata, permettono all'operatore di discriminare meglio la scelta del codice all'interno dei nove Grandi Gruppi. La tabella che segue evidenzia le *stringhe/parole chiave* che sono state contemplate nei testi delle empiriche del dizionario e i Grandi Gruppi/Gruppi all'interno dei quali esse permettono di discriminare meglio la ricerca

Tabella 6.1 –Elenco delle stringhe/parole chiave

<b>Stringhe/Parole chiave</b>	<b>Grande Gruppo/Gruppo</b>
<i>Dirigente pubblico/privato</i>	Gruppo 1.1 e 1.2
<i>Imprenditore grande/piccola azienda privata</i>	Gruppo 1.2 e 1.3
<i>Gestore/titolare</i>	Gruppo 1.3
<i>Specialista</i>	Grande Gruppo 2
<i>Tecnico/impiegato direttivo/impiegato responsabile/impiegato coordinatore</i>	Grande Gruppo 3 e Gruppo 3.3
<i>Impiegato esecutivo/impiegato di concetto/impiegato mansioni generiche</i>	Grande Gruppo 4
<i>Operaio specializzato/artigiano</i>	Grande Gruppo 6
<i>Operaio conduttore impianti industriali/operaio addetto macchinari/operaio lavorazioni in serie/operaio catena montaggio</i>	Grande Gruppo 7
<i>Operaio generico/operaio non qualificato/manovale/bracciante/operaio semplice/salariato</i>	Grande Gruppo 8
<i>Militare</i>	Grande Gruppo 9

Nella formazione sono stati pertanto invitati gli operatori ad utilizzare le *parole chiave* e quindi a riportarle nei testi delle *Professioni* dichiarate dagli intervistati prima di procedere ad effettuare la ricerca del testo all'interno del dizionario.

## 6.3 Formazione rivolta all'utilizzo dei Campi Note

Questo aspetto della formazione si è reso necessario per aiutare l'operatore a dirimere l'ambiguità della codifica di *Professioni* generiche non codificabili univocamente. Qualora infatti venga rilevata una *Professione* che non consente l'attribuzione di un codice univoco e completo, la presenza del *Campo Note* segnala all'operatore non solo che si sta rilevando una *Professione* generica ma, soprattutto, che è necessario chiedere all'intervistato ulteriori specifiche sulla *Professione* dichiarata per arrivare ad una codifica al massimo dettaglio.

La formazione quindi è stata mirata ad aiutare l'operatore a formulare domande mirate all'intervistato, avvalendosi anche dell'aiuto del *Campo Note*, in maniera tale da poter navigare nel dizionario informatizzato per arrivare ad una codifica corretta.

Analizziamo per esempio come l'operatore è stato istruito a comportarsi in caso di una risposta generica quale '*operaio edile*'

Descript	PROF_Cod	Note
<b>OPERAIO EDILE</b>	<b>6.1.2</b>	<b>Operai specializzati nell'edilizia generici</b>
OPERAIO EDILE	7.4.4	Operai conduttori macchine per l'edilizia generici
OPERAIO EDILE MANOVALE	8.6.2.1	
OPERAIO EDILE CARPENTERIA	6.1.2.3	
OPERAIO EDILE ASSISTENZA CANTIERE	6.1.2	Operai specializzati nell'edilizia generici
OPERAIO EDILE ASSISTENZA CANTIERE	7.4.4	Operai conduttori macchine per l'edilizia generici
MURATORE SPECIALIZZATO OPERAIO EDILE	6.1.2.1	
OPERAIO EDILE MANOVALE ASSISTENZA CA	8.6.2.1	
OPERAIO GRUISTA EDILE	7.4.4.3	
OPERAIO LAVORATORE EDILE	6.1.2	Operai specializzati nell'edilizia generici
OPERAIO LAVORATORE EDILE	7.4.4	Operai conduttori macchine per l'edilizia generici
OPERAIO MANOVALE LAVORATORE EDILE	8.6.2.1	
OPERAIO SPECIALIZZATO CANTIERE EDILE	6.1.2	Operai specializzati nell'edilizia generici
OPERATORE SPECIALIZZATO EDILE	6.1.2	Operai specializzati nell'edilizia generici
OPERAIO SPECIALIZZATO EDILIZIA	6.1.2	Operai specializzati nell'edilizia generici
OPERAIO IMPASTATORE MECCANICO EDILIZ	7.2.1.2	
OPERATORE MANOVALE EDILE	8.6.2.1	
OPERAIO MANOVRATORE MACCHINE EDILI	7.4.4	Operai conduttori macchine per l'edilizia generici
OPERAIO MANOVALE MANUTENZIONE EDIFI	8.6.2.1	
OPERAIO SPECIALIZZATO ADDETTO RIFINIT	6.1.3	Operai specializzati addetti alle rifiniture edili generici
OPERAIO MACCHINARI MOBILI PERFORAZIO	7.4.4.2	

Search: operaio edile

Key type:  TrigrDesc  AlphaDesc

1:299

Select Cancel

Avvalendosi della schermata sopra riportata, l'operatore avrebbe dovuto formulare le seguenti domande *mi scusi*:

- è' un operaio specializzato dell'edilizia?
- è' un operaio conduttore di macchine per l'edilizia?
- è' un operaio manovale edile?

Una volta raccolte le informazioni mancanti e riportate nel campo di ricerca *Search* il *software* avrebbe proposto una nuova schermata dove effettuare la scelta del codice completo.

Sono stati forniti, inoltre, esempi di *testi da non rilevare* in quanto privi o delle informazioni di dettaglio oppure contenenti informazioni inutili ai fini della codifica. Un esempio di questa tipologia di testi è '*amministrazione azienda privata*'. Si vede che essa risulta non codificabile univocamente in quanto chi si dichiara tale può essere:

- *amministratore di grande azienda privata* (manca ancora informazione → settore dove opera l'azienda)
- *amministratore di piccola azienda privata* (manca ancora informazione → settore dove opera l'azienda)
- *tecnico amministrazione azienda privata* (informazione inutile → azienda privata)
- *impiegato esecutivo amministrazione azienda privata* (informazione inutile → azienda privata)

#### 6.4 Formazione sugli aspetti tecnici

Coerentemente con la standardizzazione effettuata sui testi del dizionario (*cf. capitolo 5*) agli operatori CATI è stato chiesto di effettuare una rielaborazione dei testi rilevati prima di procedere con la ricerca. Gli intervistatori dovevano, cioè:

- digitare il testo senza congiunzioni, articoli, preposizioni e avverbi;
- omettere nella digitazione del testo l'accento e l'apostrofo;
- riportare le *Professioni* al '*maschile singolare*';
- rilevare il testo senza errori di ortografia;
- non usare né abbreviazioni né sigle all'interno del testo;
- cercare di non riportare testi eccessivamente lunghi.

Agli operatori è stato suggerito anche di evitare l'utilizzo di certe stringhe e parole che, in quanto prive di informatività, non solo non discriminano tra le varie *Professioni*, ma possono anche determinare falsi *match* ad esempio del tipo:

- *lavoratore /dipendente/libero professionista;*

- *produttore/produce;*

- *costruttore/costruisce/costruzioni/costruzione* usarle solo se indicano una professione all'interno dell'edilizia;

- *qualifica/mansioni/ramo/settore/ambito/mi occupo/svolgo/faccio/lavoro/fa parte;*

- *presso agenzia/azienda privata/azienda di famiglia/attività familiare/conduzione familiare/in proprio/presso privato/studio privato/sotto padrone.*

In particolare sono state individuate le casistiche comportamentali da seguire a seconda dei possibili esiti del *matching* testuale tra il testo rilevato, opportunamente standardizzato, ed i testi del dizionario elaborabile. Diverso infatti è il comportamento individuato per l'intervistatore a seconda che nell'elenco di *Professioni* proposte dal sistema *software* a seguito del *matching*:

- nella schermata di *Look-up* sia presente una *Professione* identica con un codice a quattro digit;
- nella schermata di *Look-up* non esiste una *Professione* identica a quella digitata ma ce ne siano alcune simili corrispondenti a codici a quattro digit;
- nella schermata di *Look-up* non esistano *Professioni* simili corrispondenti a codici completi.

Nel primo caso, l'operatore è stato istruito ad evidenziare il testo della *Professione* con il relativo codice dalla finestra di *Look-up* e, tramite il tasto *Select* ad effettuare la codifica premendo poi *Invio* per passare al quesito successivo

Nel secondo caso, l'operatore è stato istruito a richiedere ulteriori informazioni per discriminare tra i codici proposti e a riportare le risposte testuali nel campo di ricerca. La selezione della risposta più simile a quella digitata tra quelle della nuova schermata proposta dal *software* doveva quindi avvenire selezionando il testo e premendo *Invio*.

Nel terzo caso l'operatore è stato istruito ad effettuare ulteriori domande a seconda di quanto riportato nei *Campi Note* associati ai codici generici e a procedere quindi come nei casi precedenti.

Nel caso infine in cui, nel corso dell'intervista, la codifica della *Professione* del diplomato o laureato, fosse risultata troppo complessa oppure il codice assegnato *on line* non fosse stato completo, si aveva la possibilità di rinviare la codifica a fine intervista. In questo caso l'operatore doveva digitare tutte le informazioni necessarie per la codifica a posteriori nell'apposito campo '*professione estesa*', secondo la logica precedentemente descritta, utilizzando cioè le *parole chiave* e avendo cura di specificare le informazioni di dettaglio necessarie per la codifica completa. Nella schermata della codifica a posteriori, l'operatore avrebbe trovato poi oltre il campo con la professione estesa dichiarata dal diplomato/laureato in fase di intervista, anche altre variabili correlate con la *Professione* (es *posizione nella professione, settore...*) che potevano essergli d'ausilio per la codifica finale.

## **7 Test sull'efficacia della formazione rivolta agli Operatori CATI**

Dal momento che quella effettuata in occasione delle due Indagini sulla *Transizione istruzione-lavoro* è stata la prima esperienza di utilizzo dell'applicazione di codifica assistita, si è ritenuto opportuno effettuare un'analisi di qualità sulla rilevazione e codifica dei testi della *Professione* finalizzata a verificare l'efficacia della formazione rivolta agli operatori CATI e contemporaneamente individuare quelli che necessitavano di un ulteriore *briefing*.

L'analisi ha riguardato la coerenza tra testi rilevati e i codici assegnati ed è stata effettuata su 4105 testi di *Professioni* dichiarate dai diplomati a distanza di circa un mese dall'inizio dell'Indagine. La decisione di effettuare un'analisi a tappeto sulle risposte fornite è giustificata dal fatto che, come già detto, si trattava di una esperienza innovativa. A regime, invece, sarebbe opportuno procedere selezionando periodicamente un campione di testi da sottoporre a validazione.

Sono state contemplate nell'errore di codifica le seguenti casistiche:

- *Professioni rilevate in maniera incompleta* (senza cioè le informazioni di dettaglio ed a cui invece è stato associato un codice completo a quattro digit); ad esempio professione rilevata '*agente*', (codice assegnato 3342). La schermata che appare all'interno del modulo di codifica a seguito della digitazione di '*agente*' è la seguente:

Descript	PROF_Code	Note
AGENTE CAVA	6.1.1.5	
AGENTE BORSA	3.3.2.5	
AGENTE DAZIO	8.1.1.1	
AGENTE PEGNI	4.2.1.6	
AGENTE AFFARI	3.3.4.2	
AGENTE CAMBIO	3.3.2.5	
AGENTE SANITA	5.4.1.0	
AGENTE GIURATO	5.5.4.6	
AGENTE MANOVRA	7.4.1.2	
AGENTE MINIERE	6.1.1.5	
AGENTE POSTALE	4.1.4.7	
AGENTE VENDITA	3.3.4.6	
AGENTE VIAGGIO	3.4.1.4	
AGENTE BREVETTI	4.1.2.3	
AGENTE TEATRALE	3.3.4.9	
AGENTE COMMERCIO	3.3.4.2	
AGENTE CONTABILE	3.3.1.2	
AGENTE LOCAZIONE	3.3.4.5	
AGENTE MARITTIMO	3.3.4.1	
AGENTE NECROFORO	5.5.3.9	

Search: agente

Key type:  IgrDesc  AlphaDesc

Select Cancel

1:285

Start | Analisi\_cod... | diplomati\_2... | [Fwd: Sche... | Blaise Dat... | 15.24

Come si vede, ogni etichetta contenente la parola 'agente' presenta anche una specifica diversa del tipo *cava*, *borsa*, *affari*...; l'operatore pertanto, avendo assegnato il codice 3342, proprio degli 'agenti di commercio', o ha effettuato una forzatura oppure non ha trascritto nel campo di ricerca *Search* l'informazione di dettaglio raccolta successivamente dall'intervistato, necessaria per l'assegnazione del codice corretto.

- *Errore di selezione del codice* (la *Professione* è stata rilevata in maniera corretta ma viene effettuato un errore nella scelta del codice); ad esempio professione rilevata 'musicista' (codice assegnato 3446). La schermata che appare a seguito della digitazione di 'musicista' è la seguente:

Descript	PROF_Code	Note
<b>MUSICISTA</b>	<b>2.5.5.4</b>	
MUSICISTA BANDA	3.4.4.6	
MUSICISTA VIOLINO	2.5.5.4	
MUSICISTA CANTANTE	2.5.5.4	
MUSICISTA PIANO BAR	2.5.5.4	
MUSICISTA PIANOFORTE	2.5.5.4	
MUSICISTA CHITARRISTA	2.5.5.4	
STRUMENTISTA MUSICISTA	2.5.5.4	
MUSICISTA CONSERVATORIO	2.5.5.4	
MUSICISTA LOCALI PIAZZE	3.4.4.6	
MUSICISTA BASSO ELETTRICO	2.5.5.4	
MUSICISTA TASTIERA ORCHESTRA	2.5.5.4	
MUSICISTA CONCERTI CLARINETTO	2.5.5.4	
CLAVISTA STRUMENTI MUSICALI	6.3.1.3	
ARTIGIANO CLAVISTA STRUMENTI MUSICALI	6.3.1.3	
PSICOMETRICISTA	2.5.3.3	
PSICOMOTRICISTA	3.2.1.4	
EDUCATORE PSICOMOTRICISTA	3.4.2.2	
ELICISTA	6.2.3.1	
MANTICISTA	6.3.1.3	

Search: musicista

Key type:  TrigrDesc  AlphaDesc

Select Cancel

1:98

Start | Analisi\_cod... | diplomati\_2... | Fwd: Sche... | Blaise Dat... | 15.46

Come si vede nella schermata è presente il testo *'musicista'* associato ad un codice completo a 4 digit, diverso da quello attribuito. L'operatore, però, avendo assegnato il codice 3446, proprio dei *'musicisti di banda'*, o ha erroneamente selezionato il codice proposto nella seconda riga della schermata, oppure si è dimenticato, anche in questo caso, di trascrivere nel campo di ricerca *Search* le informazioni mancanti.

- *Codifiche fatte a posteriori dubbie.* Alla *Professione* rilevata *on line* non si è riusciti ad assegnare alcun codice oppure è stato assegnato un codice incompleto; era necessario pertanto riportare nel campo testo la *Professione* per esteso con le informazioni di dettaglio da raccogliere in fase di intervista. Sono state considerate dubbie le codifiche effettuate a posteriori assimilabili al seguente esempio:

- Testo professione rilevata on line *assistente veterinario*
- Testo professione rilevata on line nel campo esteso *assistente veterinario*
- Testo codificato a posteriori *segretario studio*
- Codice attribuito *4.1.1.4*

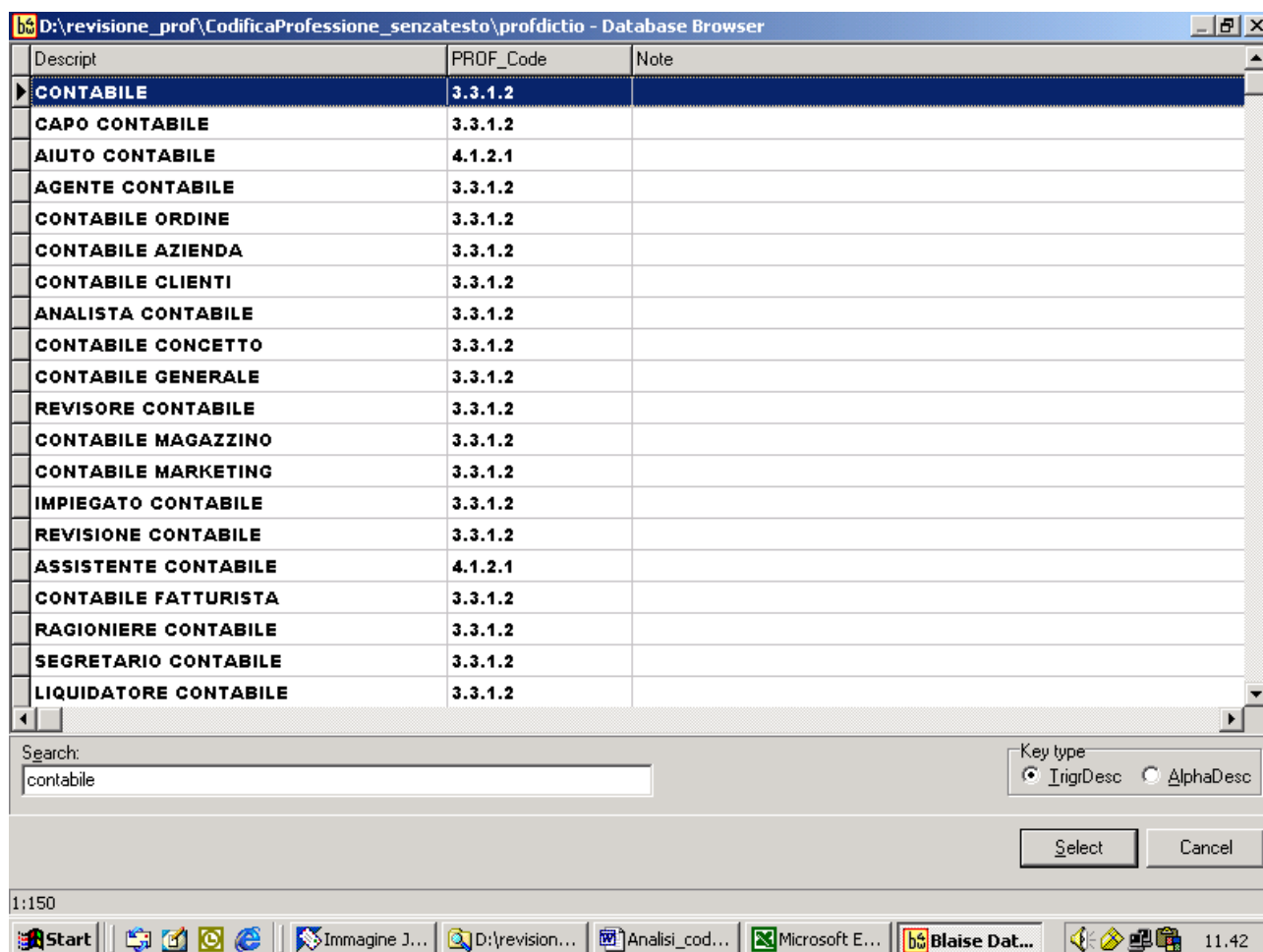
In questo caso l'operatore non solo non ha rilevato bene la *Professione* nel *campo esteso* (ha riportato lo stesso testo rilevato *on line* senza aggiungere le informazioni di dettaglio necessarie per la codifica a posteriori), ma ha anche assegnato un codice completo, la cui correttezza rimane dubbia (all'interno del dizionario informatizzato al testo riportato nel campo esteso non corrisponde il codice assegnato). A questo si può aggiungere anche che le informazioni di dettaglio necessarie per una codifica a quattro digit, non può averle dedotte neanche dalle altre

variabili presenti nella schermata della codifica a posteriori (lavoro dipendente o autonomo, posizione nella professione, settore di attività economica).

Rientra sempre in questa casistica di errore le codifiche effettuate a posteriori che potevano invece essere risolte in corso di intervista. Ad esempio

- Testo professione rilevata on line *contabile*
- Testo professione rilevata on line nel campo esteso *contabile piccola azienda*
- Testo codificato a posteriori *contabile3312*
- Codice attribuito *3.3.1.2*

La schermata che appare all'operatore in corso di intervista è infatti la seguente



Come si vede nella schermata è presente lo stesso testo di input associato ad un codice completo a 4 digit. L'operatore non ha però selezionato direttamente il codice ed ha invece riportato il testo nel campo esteso aggiungendo anche delle specifiche inutili ai fini della rilevazione e quindi della codifica.

Quantificando, in base alle casistiche d'errore sopra descritte, si sono avute in totale 280 *Professioni* (sui 4105 testi rilevati) che sono state rilevate/codificate in maniera non corretta. Si è avuta pertanto una *percentuale d'errore totale di codifica del 7%*; inoltre delle 280 *Professioni* considerate errate, 13 casi presentano un codice non completo (<4 digit).

Di contro si è avuta una *qualità di codifica pari al 93%* (*Professioni* rilevate e codificate correttamente al massimo dettaglio) come sintetizza la tabella che segue:



Tabella 7.1 – Percentuale di errore e qualità di codifica

<i>Testi professioni</i>	<i>Numero testi professioni</i>	<i>Percentuale %</i>
Testi rilevati /codificati male	280	7%
Testi rilevati/codificati correttamente	3825	93%
Totale testi rilevati	4105	100%

Dall'analisi fatta si può pertanto concludere che l'impostazione data alla formazione degli operatori CATI ha sicuramente determinato dei buoni risultati in termini sia di rilevazione della *Professione* che di codifica ed ha permesso inoltre di evidenziare gli errori più frequenti commessi dai rilevatori:

- l'operatore CATI non sempre ha riportato nel campo di ricerca *Search* il testo con le specifiche dettagliate. Ha proceduto, cioè, a selezionare il codice e il testo direttamente dalla schermata di *Look-up*, senza trascrivere le informazioni raccolte successivamente nel campo ricerca e questo ha comportato un'apparente incompatibilità tra testo e codifica effettuata;
- l'operatore CATI, pur rilevando bene la *Professione*, ha poi codificato male ovvero ha selezionato un codice non coerente con il testo rilevato, errore da imputare sicuramente alla fretta per il timore di allungare troppo i tempi dell'intervista.

Questa analisi ha dato quindi l'opportunità di effettuare un ulteriore *briefing* agli operatori in linea con le osservazioni riportate.

La fase di addestramento e di ampliamento del dizionario informatizzato non può e non deve, comunque, ritenersi conclusa, in quanto l'analisi per esempio dei testi rilevati all'interno delle due Indagini, potrebbe aiutare ad implementare ulteriormente il dizionario informatizzato con *Professioni* non ancora previste e contemporaneamente potrebbe fornire uno strumento per evidenziare nuove *Professioni* emergenti che possono risultare utili anche per un eventuale aggiornamento della Classificazione ufficiale.

## 8 Le sperimentazioni propedeutiche all'adozione del modulo della codifica assistita nelle indagini CATI sulla Transizione istruzione-lavoro

### 8.1 Il sistema di indagini sulla Transizione istruzione-lavoro

L'ISTAT conduce periodicamente tre Indagini sulla condizione occupazionale dei giovani in possesso di un titolo di studio post-obbligo (diploma di scuola secondaria superiore, diploma universitario e laurea), con lo scopo di fornire strumenti per la valutazione dell'efficacia dell'istruzione superiore sul mercato del lavoro. Si tratta di un sistema integrato di rilevazioni (Sistema di Indagini ISTAT sulla *Transizione istruzione-lavoro*) il cui obiettivo principale è consentire un'analisi comparativa del rendimento occupazionale dei titoli di studio superiori.

Le Indagini sono fortemente coerenti dal punto di vista della metodologia adottata: hanno tutte cadenza triennale, vengono condotte su singole leve di diplomati (di scuola secondaria superiore, universitari o con laurea) a circa tre anni dal conseguimento del titolo e sono realizzate con tecnica CATI. L'Indagine sui diplomati universitari è totale, mentre quelle sui diplomati di scuola secondaria superiore e sui laureati sono campionarie.

Tabella 8.1 – Caratteristiche del Sistema di Indagini Istat sulla Transizione istruzione-lavoro

	Tipo	Prima indagine		Ultima indagine	
		Leva	Rilevazione	Leva	Rilevazione
Indagini sui diplomati delle secondarie	Campionaria	1995	1998	2001	2004 (III edizione)
Indagini sui diplomati universitari	Totale	1996	1999	1999	2002 (II edizione)
Indagini sui laureati	Campionaria	1986	1989	2001	2004 (VI edizione)

L'indagine di più lunga tradizione è quella sull'*Inserimento professionale dei laureati* (ad oggi è stata conclusa la VI edizione), mentre quella di più recente realizzazione è la rilevazione sull'*Inserimento professionale dei diplomati universitari* della quale, nel 2002, è stata portata a termine la seconda ed ultima edizione.<sup>2</sup>

L'integrabilità delle Indagini è evidente anche da un punto di vista contenutistico: le rilevazioni utilizzano questionari analoghi nella struttura e, fin dove possibile, nei contenuti informativi. In sintesi, gli argomenti affrontati riguardano:

- i *percorsi formativi* (scolastici, universitari e di formazione professionale): l'obiettivo è di verificare quale sia il rapporto tra gli studi conclusi e le scelte successive al conseguimento del titolo post-obbligo, siano esse di studio o di lavoro;
- la *situazione lavorativa*: nello specifico si analizza il lavoro attuale svolto dal giovane, la sua storia lavorativa e l'eventuale ricerca di lavoro;
- le *notizie sulla famiglia*: lo scopo è di ottenere una descrizione dell'ambiente socio-culturale di provenienza degli intervistati attraverso la raccolta di informazioni sulla struttura familiare, sul livello di istruzione dei genitori e sulla loro condizione professionale;
- le *notizie anagrafiche*.

<sup>2</sup> La riforma dei cicli universitari dell'a.a. 2001/2002 (avviata in via sperimentale nell'a.a. precedente), ha, infatti, trasformato radicalmente l'organizzazione della didattica del sistema universitario, determinando tra l'altro la chiusura delle immatricolazioni ai corsi di diploma universitario. Precisamente, l'istruzione universitaria, tradizionalmente articolata in due cicli paralleli (corsi di diploma universitario e scuole dirette a fini speciali di durata biennale o triennale e corsi di laurea di durata variabile tra i 4 e i 6 anni), è ora strutturata in corsi di studio disposti su due livelli in serie: corsi di laurea di primo livello (di durata triennale), che rilasciano il nuovo diploma di laurea, corsi di laurea di secondo livello (di durata biennale), finalizzati al conseguimento della laurea specialistica e un circoscritto numero di corsi di laurea a ciclo unico (di durata pari a 5 o 6 anni) che rilasciano un titolo equipollente alla laurea specialistica di secondo livello.

La principale sezione dei questionari indaga sul lavoro svolto dall'intervistato, rilevando - in modo uniforme - caratteristiche quali la posizione nella *Professione*, il settore di attività, il tipo di contratto, il reddito, i requisiti di accesso alla *Professione* e la *Professione*. La situazione lavorativa dei genitori viene invece rilevata nella sezione sul *background* familiare. La variabile *Professione* - in particolare - viene rilevata sia per il rispondente sia per il padre. L'analisi della *Professione* del rispondente è estremamente importante per comparare le *performance* sul mercato del lavoro di titoli di studio differenti sia per livello, sia per settore disciplinare. La *Professione* paterna rappresenta, invece, accanto alle altre variabili sul *background* socio-culturale, lo strumento attraverso cui valutare l'influsso dell'estrazione sociale sulle carriere di studio e sulla riuscita professionale giovanile.

Le ultime edizioni delle *Indagini sui diplomati di scuola secondaria superiore e sui laureati del 2001* (concluse a dicembre 2004), presentano numerose novità. La principale, per quanto concerne *l'Indagine sull'inserimento professionale dei laureati*, riguarda l'ampliamento del campione a circa 26.000 unità, reso necessario per garantire risultati attendibili a livello di singolo ateneo incrociato per area didattica e permettere più puntuali valutazioni sull'efficacia occupazionale della formazione universitaria. *Nell'Indagine sui percorsi di studio e di lavoro dei diplomati di scuola secondaria superiore*, l'innovazione principale consiste - anche a seguito di una convenzione siglata con il Ministero del lavoro e delle politiche sociali - nell'aggiornamento ed ampliamento della sezione del questionario relativa alle attività di formazione professionale post-diploma. Come è infatti noto, anche in relazione ai cambiamenti introdotti con la legge quadro 53/2003 nel sistema educativo nazionale, il segmento della formazione professionale va acquisendo sempre maggiore rilevanza nei percorsi formativi dei giovani.

Innovazioni congiunte - finalizzate al miglioramento della qualità dei dati - riguardano invece:

- la ristrutturazione dei questionari allo scopo di recepire le innovazioni del mercato del lavoro e migliorare l'integrabilità delle Indagini con le altre fonti ISTAT sull'occupazione (in particolar modo la *Rilevazione Continua delle Forze di Lavoro*);
- lo sviluppo *in-house* del software funzionale alla fase di acquisizione ed il conseguente utilizzo del sistema BLAISE per l'implementazione del questionario elettronico CATI (entrambe le Indagini, infatti, nelle precedenti edizioni, erano state condotte sempre con la tecnica CATI, ma tutte le attività inerenti la fase di acquisizione erano state affidate ad una ditta esterna);
- l'adozione di un sistema di codifica assistita della *Professione*.

## 8.2 Come si è pervenuti all'adozione del modulo di codifica assistita?

Si è visto come nelle Indagini ISTAT la rilevazione della variabile *Professione* sia condizionata da norme di Classificazione che, aderendo a convenzioni internazionali, assicurano la confrontabilità dei dati, oltre che tra le varie fonti dell'Istituto, con le Indagini effettuate dagli organismi competenti in materia negli altri Paesi. Anche nel Sistema di Indagini sulla *Transizione istruzione-lavoro*, quindi, viene adottata la Classificazione ufficiale dell'Istituto (Classificazione delle professioni - CP2001).

La complessità della Classificazione, che si compone di 6.300 Voci professionali, articolate secondo un sistema gerarchico, (*cf. capitolo 4*) è tale da rendere assai difficile per il rispondente la collocazione della propria *Professione* in una delle Categorie previste dalla Classificazione ufficiale. Nelle Indagini ISTAT sulla transizione, quindi, sino ad oggi si è proceduto raccogliendo le descrizioni alfabetiche delle *Professioni*, per ricondurle ex-post alla Classificazione ufficiale. La codifica delle risposte testuali avveniva, in particolare, con l'ausilio di codificatori che procedevano "manualmente", consultando cioè il manuale della Classificazione e selezionando il codice che

ritenevano più corretto rispetto alla descrizione offerta dal rispondente.

Per le recenti tornate di *Indagine sui laureati e sui diplomati di scuola secondaria superiore del 2001*, si è deciso invece di adottare una procedura di codifica nel corso dell'intervista telefonica, facendo ricorso al modulo di codifica assistita implementato in BLAISE. La decisione è stata assunta a seguito delle positive *performance* registrate da BLAISE nell'ambito di due diverse fasi propedeutiche di sperimentazioni condotte:

- sui dati dell'*Indagine 2001 sull'Inserimento professionale dei laureati del 1998*: in questo caso le *Professioni* raccolte sono state codificate ex-post da codificatori esperti attraverso il modulo BLAISE;
- in fase di Pre-test del questionario di rilevazione messo a punto per la realizzazione dell'*Indagine 2004 sull'inserimento professionale dei laureati*: la codifica della *Professione* è avvenuta in corso di intervista.

### 8.3 Prima fase di sperimentazione: codifica a posteriori con il modulo BLAISE

La prima fase di sperimentazione è stata effettuata con due principali obiettivi:

- valutare l'applicabilità del modulo BLAISE all'Indagine in questione;
- comparare le *performance* di due metodi alternativi di codifica.

Il confronto è stato operato, in particolare, tra il metodo "*testuale*" (secondo la dizione alfabetica) consistente in una ricerca testuale per trigrammi, direttamente nel dizionario, sulla base del testo digitato - ed il metodo "*semi-gerarchico*" o "*misto*" (cfr. capitolo 3). Quest'ultimo deriva da una combinazione del metodo gerarchico e del metodo alfabetico tale per cui, in una prima fase si seleziona il Grande Gruppo corrispondente al primo digit della Classificazione<sup>3</sup> e successivamente si procede ad una ricerca testuale all'interno del Grande Gruppo e alla selezione, tra i vari codici proposti dal programma, del codice a 4 digit corrispondente alla stringa ricercata.

Un primo test è stato implementato su 200 descrizioni alfabetiche (100 per i laureati e 100 per i padri), corrispondenti alle *Professioni* più frequenti nel campione delle *Professioni*, rispettivamente, dei figli e dei padri, raccolte nell'*Indagine del 2001 sull'inserimento professionale dei laureati*. Ognuna delle 200 descrizioni alfabetiche è stata codificata ex-post sia con il metodo testuale, sia con quello semi-gerarchico.

I due metodi sono stati comparati con riferimento:

- al tempo necessario per la codifica;
- all'accuratezza dei risultati (valutata da codificatori esperti).

Per evitare il rischio di effetti codificatore, i codificatori hanno proceduto codificando parte delle *Professioni* dei laureati e parte di quelle dei padri ed utilizzando in parte il primo metodo ed in parte il secondo.

I risultati del test hanno condotto alle seguenti considerazioni:

- entrambi i metodi hanno mostrato un buon rendimento in termini di tempo necessario alla codifica (vedi tabella 8.2): il tempo medio necessario a codificare una descrizione è stato pari a 9,7 secondi per il metodo testuale (5'' minimo; 55'' massimo) e 14,9 secondi per il metodo semi-gerarchico (5'' minimo; 70'' massimo). Un tempo medio di codifica comunque inferiore ai 15 secondi – sia con il primo sia con il secondo metodo – ha incoraggiato circa la praticabilità della codifica assistita in corso di intervista. Infatti, anche considerando l'elevata probabilità di

---

<sup>3</sup> Legislatori, Dirigenti e imprenditori = 1; Professioni intellettuali, scientifiche e di elevata specializzazione = 2; Professioni tecniche = 3; Impiegati = 4; Professioni qualificate nelle attività commerciali e nei servizi = 5; Artigiani, operai specializzati e agricoltori = 6; Conduttori di impianti e operai semi-qualificati addetti a macchinari fissi e mobili = 7; Professioni non qualificate = 8; Forze Armate = 9.

tempi medi di codifica *on-line* superiori, per effetto dell'interazione tra intervistatore e intervistato, il rischio di un incremento di interruzione di intervista dovute all'eccessiva durata è sembrato nel complesso circoscritto;

- in termini di accuratezza (vedi tabella 8.3), i due metodi hanno prodotto codici corretti e coincidenti nell'83,0% dei casi, se si considerano i codici completi (a 4 digit) e nell'86,0% dei casi, assumendo come coincidenti i codici con almeno i primi 3 digit uguali;
- la procedura alfabetica ha mostrato un migliore rendimento di quella mista, non solo per quel che concerne la durata media di codifica (vedi tabella 8.2), ma anche in termini di accuratezza. Infatti, benché nella maggior parte dei casi i due metodi abbiano prodotto lo stesso risultato, un'analisi dei codici non coincidenti ha mostrato come, in 10 casi su 34 (9 su 28, se si prendono in considerazione solo i primi 3 digit) i codici corretti siano stati selezionati dal metodo alfabetico (1 solo caso per il metodo semi-gerarchico);
- durante questa prima fase di sperimentazione si è avuto, inoltre, modo di apprezzare la maggiore facilità di navigazione nel dizionario delle *Professioni* adottando il metodo testuale piuttosto che il metodo semi-gerarchico. Infatti, si è potuto constatare che con il primo metodo non è necessaria una conoscenza puntuale dalla Classificazione delle *Professioni*, indispensabile, invece con il secondo. Con l'utilizzo del metodo misto, infatti, può succedere che l'iniziale errata individuazione del Grande Gruppo (del 1° digit del codice), che potrebbe essere determinata da rilevatori poco esperti o da *Professioni* rilevate in modo generico, comprometta definitivamente gli esiti della successiva ricerca testuale.

Tabella 8.2 - Tempi di codifica della Professione con il metodo testuale e con il metodo semi-gerarchico: primo test (in secondi)

	Minimo	Medio	Massimo
<b>Metodo testuale</b>	5	9,7	55
<b>Metodo semi-gerarchico</b>	5	14,9	70

Tabella 8.3 - Le codifiche con il metodo testuale e con il metodo semi-gerarchico: primo test

	Codici a 4 digit		Codici a 3 digit	
<b>Codifiche coincidenti</b>	166	83,0%	172	86,0%
<b>Codifiche non coincidenti di cui:</b>	34	17,0%	28	14,0%
<i>Corretta con il metodo testuale</i>	10		9	
<i>Corretta con il metodo semi-gerarchico</i>	1		1	
<i>Testo ambiguo (possibili più codici)</i>	21		17	
<i>Testo insufficiente</i>	2		1	
<b>Totale professioni</b>	200	100%	200	100%

La positiva *performance* registrata dal modulo di codifica assistita di BLAISE ha incoraggiato ad una verifica dei risultati attraverso un secondo test di codifica a posteriori, condotto questa volta utilizzando unicamente il metodo testuale, su ulteriori 600 professioni (300 del laureato e 300 del padre) sempre rilevate nell'*Indagine del 2001 sull'inserimento professionale dei laureati*.

I risultati del test hanno confermato ancora una volta contenuti tempi di codifica ed un elevato tasso di successo del metodo alfabetico. In particolare, la durata media della codifica assistita è stata pari a 10,7 secondi: 10,1 secondi per la *Professione* del laureato e 11,3 per quella del padre (vedi tabella 8.4).

Tabella 8.4 - Tempi di codifica della Professione con il metodo testuale: secondo test (in secondi)

	Minimo	Medio	Massimo
<b>Totale</b>	<b>5</b>	<b>10,7</b>	<b>120</b>
Laureato	5	10,1	40
Padre	5	11,3	120

L'83,7% delle descrizioni sono state codificate correttamente, sia per le *Professioni* del laureato sia per quelle del padre (vedi tabella 8.5), mentre il restante 16,3% non sono state codificate per via di due tipi di problema:

- presenza di descrizioni alfabetiche ambigue o insufficienti (nel 13,3% dei casi);
- lacune del dizionario informatizzato associato alla Classificazione (3,0%).

Tabella 8.5 - Le codifiche con il metodo testuale: secondo test

	Totale	Laureato	Padre
<b>Professioni codificate</b>	83,7%	83,7%	83,7%
<b>Professioni non codificate per:</b>			
Testo ambiguo o insufficiente	13,3%	11,7%	15,0%
Testo non presente nel dizionario	3,0%	4,7%	1,3%
Totale	100%	100%	100%

L'analisi dettagliata dei *match* falliti ha evidenziato, più in particolare, problematiche diverse a seconda che si considerasse la *Professione* del laureato o quella del padre. Infatti, in quest'ultimo caso il tasso di insuccesso è stato più spesso alimentato dalla presenza di *Professioni* rilevate in maniera ambigua o insufficiente (15% in totale) che rendevano impossibile la distinzione tra il Grande Gruppo dei Tecnici e quello degli Impiegati (1° digit del codice uguale a 3 o 4) oppure tra il gruppo degli Imprenditori, gestori e responsabili di piccole imprese (1° e 2° digit uguale 1 e 3) e il Grande Gruppo delle Professioni qualificate nelle attività commerciali e nei servizi (1° digit del codice uguale a 5).<sup>4</sup>

Nel caso, invece, del laureato l'ambiguità/insufficienza dei testi - un fenomeno comunque più circoscritto rispetto alle *Professioni* del genitore - rendeva difficile la distinzione tra il Grande Gruppo delle Professioni intellettuali, scientifiche e di elevata specializzazione e quello dei Tecnici (1° digit del codice uguale a 2 o 3).<sup>5</sup>

L'incidenza delle mancate codifiche delle *Professioni* rilevate per i giovani laureati è stata, comunque, spesso influenzata anche dall'assenza nel dizionario informatizzato, in taluni casi, di *Professioni* connesse al conseguimento di specifiche lauree (ad esempio l'*ingegnere gestionale*), in altri di sinonimi di più ampio utilizzo relativi a *Professioni* presenti nel dizionario con altre descrizioni (ad esempio l'*ingegnere informatico* e l'*analista di software* presenti nella Classificazione come *ingegnere software* o *analista di programmi* o *analista di sistemi*).

<sup>4</sup> Un caso tipico di Professione ambigua riscontrata per il padre è stata quella di "Bancario", codificabile come:  
 - un tecnico del lavoro bancario, ossia un operatore crediti/titoli (con codice completo uguale a 3.3.2.2);  
 - un funzionario di banca (con codice completo uguale a 3.3.2.9);  
 - un addetto allo sportello bancario, ovvero un cassiere di banca (con codice completo uguale a 4.2.1.2).

<sup>5</sup> Un esempio concreto è dato dalla descrizione di "Consulente ambientale", codificabile come:  
 - un architetto specializzato nel recupero ambientale (con codice completo uguale a 2.2.2.0);  
 - un tecnico ambientale (con codice completo uguale a 3.1.5.3).

#### 8.4 Seconda fase di sperimentazione: codifica on-line

I positivi esiti della prima fase di sperimentazione ex-post del modulo di codifica assistita hanno fatto propendere, dunque, per l'adozione del modulo BLAISE con ricerca testuale della *Professione* nelle *Indagini 2004 sui diplomati di scuola secondaria superiore e sui laureati del 2001*.

A questa fase della sperimentazione è stato necessario, però, farne seguire un'altra che testasse il processo di codifica *in corso di intervista*. Infatti, se da un lato la codifica assistita ha il vantaggio di sfruttare l'interazione tra intervistato ed intervistatore, dall'altro se non tenuta sotto controllo, può dare origine ad eccessivi allungamenti dei tempi di intervista, con il rischio di incrementare il tasso di interruzione delle Indagini. Il beneficio di una più puntuale e corretta codifica del procedimento assistito rispetto a quello manuale, grazie alla possibilità sia di richiedere ulteriori informazioni all'intervistato e sia di lasciar decidere direttamente quest'ultimo, sull'effettiva corrispondenza tra *Professione* dichiarata e *Professione* selezionata, rischia quindi di essere vanificato.

Per questo motivo è stato deciso di effettuare un nuovo test che analizzasse l'impatto della codifica *on-line* sulla:

- durata dell'intervista;
- durata della codifica della *Professione*.

Ed è proprio durante la realizzazione del primo pre-test del nuovo questionario di rilevazione per l'*Indagine 2004 sull'inserimento professionale dei laureati*, necessario per testarne la struttura, il *wording* e i nuovi quesiti, che è stato possibile analizzare l'effettivo impatto della codifica assistita sui tempi di intervista.

Realizzato alcuni mesi prima dell'avvio delle Indagini, il pre-test è stato messo a punto nel laboratorio CASIC (*Computer Assisted Survey Information Collection*) dell'Istituto, in cui sono state portate a termine, da parte di ricercatori dell'ISTAT esperti dell'Indagine in questione, 50 interviste a giovani laureati.

Ancora una volta i risultati ottenuti sono stati valutati positivamente: l'intervista di pre-test ha avuto una durata media di 19,2 minuti, contro i 18,8 registrati nel primo giorno di rilevazione nella precedente Indagine (2001) (vedi tabella 8.6).<sup>6</sup> Dei circa 19 minuti, poco più di 1 (74 secondi) è stato impiegato per codificare la *Professione* del laureato: si è dunque registrato un incremento di circa 64 secondi rispetto il tempo di codifica registrato nel secondo test realizzato ex-post (*cf. paragrafo 8.3*), perlopiù imputabile alla interazione tra intervistato ed intervistatore.

Tabella 8.6 – Durata dell'intervista e della codifica della *Professione* nel primo pre-test 2004 e nell'Indagine 2001 sull'inserimento professionale dei laureati (in minuti)

	PRE-TEST 2004		INDAGINE 2001	
	Intervista	Codifica professione del laureato	Intervista (1° giorno di rilevazione)	Intervista (Totale)
Media	19,2	1,2	18,8	10,9

L'ottimismo per le contenute durate sia dell'intervista e sia della codifica della *Professione* è stato però attenuato dalla preoccupazione emersa in fase di analisi dei tempi massimi di codifica: dai 4 minuti si è arrivati a registrare una durata massima di circa 7 minuti di codifica della *Professione* del laureato. Pur consapevoli che i casi di durate così lunghe sarebbero stati piuttosto circoscritti, si è ritenuto necessario riflettere sulla possibilità di limitare il tempo che l'intervistatore poteva dedicare alla codifica *on-line*, onde evitare l'innalzamento del tasso di interruzione.

<sup>6</sup> È importante confrontare i tempi medi di intervista di rilevatori che, approssimativamente, abbiano realizzato lo stesso numero di interviste, poiché, come sarà possibile analizzare in seguito, la durata pro-capite dell'intervista diminuisce notevolmente all'aumentare dell'esperienza.

È stato deciso, quindi, di realizzare un secondo pre-test avente come unico obiettivo quello di esaminare i tempi di codifica della *Professione*. Sono state testate, quindi, solo le domande del questionario inerenti il lavoro svolto sottoponendole a 63 laureati nell'ambito di nuove interviste realizzate sempre nel laboratorio CASIC dell'Istituto.

Tabella 8.7 – Durata della codifica della *Professione* del laureato e del padre: secondo pre-test (in minuti)

	Laureato	Padre
Minima	0,3	0,2
<b>Media</b>	<b>1,1</b>	<b>0,8</b>
Massima	2,8	3,5

Il tempo medio di codifica della *Professione* del laureato è rimasto perlopiù identico (1,1 minuto contro 1,2 del primo pre-test), mentre il tempo massimo registrato è stato più basso (2,8 minuti contro 6,9). La durata media della codifica della *Professione* del padre, invece, è stata pari a poco meno di un minuto (48 secondi), contro un tempo massimo di 3,5 minuti (210 secondi).

L'analisi della distribuzione delle interviste in base ai tempi di codifica della *Professione* (vedi tabella 8.8) ha inoltre evidenziato che:

- il 50% delle *Professioni* sono state codificate, per il laureato, in circa 1 minuto, per il padre, in poco più di ½ minuto;
- la quasi totalità (90%) delle *Professioni* sono state codificate in poco più di 2 minuti, per il figlio, in quasi 1 minuto e ½ per il padre.

Tabella 8.8 – Distribuzione dei tempi di codifica della *Professione* del laureato e del padre: secondo pre-test (in minuti)

	Laureato	Padre
25 %	< 0,5	< 0,4
50 %	< 0,9	< 0,6
75 %	< 1,4	< 1,1
90 %	< 2,1	< 1,4

Sebbene più contenuti rispetto a quelli riscontrati nel precedente pre-test, anche in questo secondo test i tempi massimi di codifica sono risultati piuttosto elevati, tanto da optare - nella soluzione da adottare per l'Indagine definitiva - per l'introduzione di una limitazione temporale alle operazioni di codifica. Per contenere il più possibile il rischio di interruzione di intervista, si è dunque deciso di introdurre un messaggio a video che consigliasse al rilevatore di procedere con l'intervista nel caso in cui, trascorso un certo lasso temporale dal momento di avvio della navigazione nel dizionario della Classificazione, non fosse riuscito ad individuare un codice completo a 4 digit. Sulla base dei risultati registrati nel pre-test, tale lasso temporale è stato fissato a 2 minuti, il tempo ritenuto sufficiente a codificare almeno il 90% delle *Professioni* sia dei laureati sia dei padri (vedi tabella 8.8).

Nel caso in cui la durata della codifica ecceda i 2 minuti, la schermata CATI che appare all'intervistatore è la seguente:



Codice

Descript	PROF_Code	Note
CONSULENTE AZIENDALE	2.5.1.3	
CONSULENTE GESTIONE AZIENDALE	2.5.3.1	
CONSULENTE ORGANIZZAZIONE AZIENDALE	2.5.3.1	
CONSULENTE DIREZIONE ORGANIZZAZIONE AZIENDALE	3.3.1.1	
CONSULENTE SICUREZZA PATRIMONIO AZIENDALE	3.3.2.1	
IMPIEGATA BANCARIA CONSULENTE COMMERCIALE CLIENTELA	3.3.2.2	
CONSULENTE LEGALE	2.5.2.2	
CONSULENTE FISCALE	2.5.1.4	
CONSULENTE MUSICALE	2.5.5.1	
CONSULENTE SINDACALE		
CONSULENTE COMMERCIALE		
CONSULENTE LAVORO FISCALE		
CONSULENTE CORSI ORIENTAM		
CONSULENTE FISCALE LIBERO		
CONSULENTE AMMINISTRATIV		
CONSULENTE TECNICO INFOR	3.3.2.4	
SPECIALISTA RECUPERO CONSERVAZIONE AMBIENTE	2.2.2.0	
CONSULENTE COMMERCIALE LIBERO PROFESSIONISTA	2.5.1.5	
CONSULENTE AMMINISTRATIVO LIBERO PROFESSIONISTA	2.5.1.2	
CONSULENTE PROPRIETA' INDUSTRIALE BREVETTI MARCHI	2.5.2.2	
CONSULENTE AMMINISTRATIVO PUBBLICA AMMINISTRAZIONE	2.5.1.1	
MONTATORE CABINE ELETTRONICHE CONTROLLO AMBIENTALE	6.2.4.1	
CONSULENTE PREVIDENZIALE PRODOTTI ASSICURATIVI FINANZIARI	3.3.2.3	

Search: consulente ambientale

Key type:  TrigrDesc  AlphaDesc

1:73

Select Cancel

Start Blaise Data Entry Documento1 - Microsoft ... IT 13.15

**ATTENZIONE** hai superato i 2 minuti: procedere con l'intervista  
ok

È importante sottolineare che si tratta di un suggerimento - e non di un'imposizione - che avverte il rilevatore che è trascorso un lasso di tempo oltre il quale si corre il pericolo di affaticare eccessivamente l'intervistato, pur lasciando libero l'intervistatore stesso di concludere le operazioni di codifica, qualora sia vicino al risultato finale.<sup>7</sup> Nel caso in cui, invece, l'intervistatore accetti il consiglio e proceda nell'intervista senza portare a termine la codifica *on-line*, limitatamente alle *Professioni* del laureato e del diplomato (non dei rispettivi padri), il programma CATI passa ad una schermata in cui l'intervistatore raccoglie tutte le informazioni possibili sui compiti esattamente svolti dal laureato/diplomato<sup>8</sup>. Tali informazioni vengono utilizzate per le operazioni di codifica che l'intervistatore, con l'ausilio del manuale della Classificazione delle *Professioni*, porta a termine a fine intervista (*cfr. paragrafo 3.1*).

<sup>7</sup> Vista la decisione di non imporre al rilevatore di proseguire con l'intervista nel caso in cui, trascorsi due minuti, non sia riuscito nella codifica della professione, si è cercato, durante la fase di formazione degli intervistatori coinvolti nella indagine definitiva, di sensibilizzarli particolarmente al pericolo di elevati tassi di interruzione, i cui valori - nell'indagine svolta nel 2004 (si veda l'ultimo capitolo) - sono risultati infatti molto bassi.

<sup>8</sup> In questo caso la domanda che viene posta al laureato/diplomato è: "Quali sono esattamente i compiti principali che svolge nel suo lavoro?"

## 8.5 Conclusioni

I risultati delle diverse sperimentazioni di codifica della *Professione*, sia di tipo ex-post e sia di tipo *on-line*, realizzate nell'ambito dei lavori preparatori dell'*Indagine 2004 sull'inserimento professionale dei laureati*, hanno condotto alle seguenti decisioni:

- adozione della codifica assistita, ritenuta non troppo onerosa in termini di tempo necessario;
- adozione del metodo testuale in quanto migliore metodo di navigazione all'interno del dizionario della Classificazione in termini di tempo di codifica, di tasso di successo, nonché ritenuto di più facile utilizzo;
- introduzione del limite temporale alle operazioni di codifica.

Infine, l'analisi dettagliata dei *match* falliti in fase di sperimentazione – a causa della presenza di testi ambigui o insufficienti nelle *Professioni* rilevate o di lacune del dizionario associato alla Classificazione – ha fornito diversi spunti per il miglioramento del processo complessivo ai fini della realizzazione della rilevazione definitiva. Nel corso dell'*Indagine 2004*, infatti, il *training* degli intervistatori, in questo senso orientato dall'esperienza svolta, ha posto particolare attenzione alla formazione alla raccolta di descrizioni alfabetiche adatte (non ridondanti, né insufficienti o ambigue) all'utilizzo della procedura assistita di codifica con metodo alfabetico (*cf. capitolo 6*). La rilevazione, inoltre, è stata preceduta da un accurato lavoro di arricchimento del dizionario informatizzato della Classificazione delle *Professioni*. Infatti, i dizionari dei manuali ufficiali il più delle volte non sono "tarati" sul *target* della specifica *Indagine*, per cui, per garantire il maggior numero di abbinamenti (e quindi di codifiche) è stato necessario introdurre delle occorrenze che, alcune volte, erano sinonimi di *Professioni* già presenti nel dizionario ed, altre, riguardavano tipiche *Professioni* dei laureati non presenti nell'elenco ufficiale della Classificazione. L'obiettivo principale dell'ampliamento è stato quello di avvicinarsi il più possibile al linguaggio del rispondente ed è per tale motivo che per realizzarlo sono state utilizzate sia le descrizioni raccolte nelle precedenti indagini del 2001 sulla *Transizione istruzione-lavoro*, sia quelle emerse in fase di pre-test dei questionari di rilevazione.

In particolare, le *Professioni* aggiunte al dizionario possono essere ricondotte alle seguenti tre tipologie:

- **Professioni espresse in inglese.** Tra le più frequenti svolte dai giovani laureati italiani sono da segnalare: l'*Account junior manager* (codice 3336), lo *Specialist account manager* (codice 2515), l'*Auditor* (codice 3312), l'*Auditor manager* (codice 2512), l'*Assistant informer front office* (codice 5134), *Junior controller* (codice 3313).
- **Professioni emergenti.** Alcuni esempi sono dati da: *Tecnico controllo di gestione* (codice 3313), *Tecnico controllo direzionale* (codice 3313), *Tecnico assistente clientela* (codice 3334), *Formatore didattico* (codice 2654), *Esperto selezione personale* (codice 2513).
- **Professioni legate al conseguimento di determinati titoli di studio.** Esempi appartenenti a tale tipologia sono: l'*Analista software* (codice 2114), l'*Ingegnere ambientale* (codice 2219), il *Consulente strategico* (codice 2531).

## 9 Il monitoraggio della codifica assistita della *Professione* nelle indagini CATI attraverso le carte di controllo

La complessità del processo di codifica assistita della *Professione* e la necessità di tenere sotto controllo l'impatto di tale complessità sulla qualità dei dati rilevati, hanno stimolato la sperimentazione di nuovi strumenti per il monitoraggio della codifica assistita, inizialmente da affiancare e successivamente da sostituire alla tradizionale reportistica (cfr. paragrafo 9.2).

Lo strumento metodologico che è stato adottato è quello delle carte di controllo. Le carte di controllo sono state già applicate con successo all'Istat nell'Indagine CATI sulla *Sicurezza del cittadino* nel 2002 (Grassia, Simeoni, in corso di stampa). In tal caso, le carte sono state applicate al monitoraggio del comportamento degli intervistatori in fase di contatto con le famiglie e venivano quindi utilizzate per monitorare gli esiti delle chiamate (tassi di rifiuto, mancato contatto...). In questa occasione, tale esperienza non è stata ripetuta sia perché per il monitoraggio della fase di contatto erano già stati predisposti numerosi *report ad hoc* e sia perché si è voluto sperimentare l'efficacia dello strumento anche per il monitoraggio di una variabile complessa quale la codifica della *Professione*.

L'ambito di applicazione del nuovo strumento è quello delle Indagini CATI, che risultano le più adatte all'uso delle carte di controllo, sia per la disponibilità immediata dei dati, sia perché l'assegnazione dei numeri telefonici agli intervistatori avviene in maniera casuale; è ed su questa casualità che si basano le ipotesi per la costruzione e applicazione delle carte di controllo.

### 9.1 Le carte di controllo

Le metodologie per il controllo statistico della qualità di processo, tra cui le carte di controllo, sono state sviluppate e trovano larga applicazione nel campo dei processi di produzione industriale ormai da decenni. La loro applicazione ai processi produttivi statistici è invece molto meno diffusa, sebbene caldamente raccomandata in letteratura (Lyberg et al., 1997; Jones, Lewis, 2004). In particolare, non risultano precedenti applicazioni delle carte di controllo al monitoraggio del processo di codifica assistita.

Le carte di controllo sono uno strumento grafico di controllo di processo che permette di discernere le fonti di variabilità di un processo e fornisce indicazioni sul modo più adatto di intervenire per migliorare la qualità del processo stesso. La variabilità presente, e mai completamente eliminabile, in un processo può essere scomposta in variabilità dovuta a cause comuni e variabilità dovuta a cause speciali (WESTAT, 1992). Le cause speciali sono avvenimenti straordinari che accrescono la normale variabilità insita nel processo, che viene a sua volta attribuita alle cause comuni. Se sono presenti cause speciali, il processo è detto fuori controllo statistico. Poiché le cause speciali sono avvenimenti particolari, in genere, per eliminare la variabilità dovuta ad esse è sufficiente intervenire con azioni circoscritte. Se sono presenti solo cause comuni di variabilità, il processo è detto sotto controllo statistico, o stabile; in questo ultimo caso, però, la variabilità presente (cause comuni) potrebbe essere ancora troppo elevata rispetto a quella desiderata e per ridurla è necessario operare cambiamenti radicali nel processo. Una volta che il processo è stabile e la sua variabilità è uguale o inferiore a quella desiderata, il processo si dice aderente alle specifiche, o capace (traduzione della parola inglese "*capable*").

Con le carte di controllo vengono monitorate le variabili chiave del processo, che sono degli indicatori, spesso indiretti, dal cui andamento si traggono informazioni sulla qualità del processo stesso: ad esempio, se si vuole monitorare il comportamento degli intervistatori nel contatto con la famiglia, il tasso di rifiuto per intervistatore rappresenta sicuramente una variabile chiave. Per applicare le carte di controllo è necessario aver preventivamente identificato le variabili chiave.

Nelle carte di controllo viene quindi monitorato l'andamento delle variabili chiave del processo con la loro variabilità osservata e, tramite regole standard, è possibile stabilire se questa è dovuta a cause speciali o a cause comuni.

Tra le diverse tipologie esistenti di carte di controllo, in questa occasione sono state utilizzate quelle di tipo Shewhart. Tra le carte di controllo di tipo Shewhart, inoltre, sono state utilizzate le  $\bar{X}$ -chart e le p-chart. Le  $\bar{X}$ -chart fanno parte del gruppo delle carte di controllo dette "per variabili" e si usano per monitorare la media di variabili continue; le p-chart fanno parte delle carte di controllo per attributi e si utilizzano per monitorare il parametro p di variabili binomiali.

Per comprendere più facilmente la costruzione di una carta di controllo si procede con un esempio (fig. 9.1). Anche se non è stato ancora descritto quali siano le variabili chiave individuate, è abbastanza intuitivo che il tempo impiegato dall'intervistatore per codificare la *Professione* lo sia. Se è troppo elevato può indicare difficoltà dell'intervistatore a codificare e può provocare interruzioni di intervista, se è troppo basso può evidenziare superficialità nell'assegnare il codice. Si suppone quindi di voler costruire una carta di controllo di tipo  $\bar{X}$ -chart per monitorare l'andamento nel tempo di un certo intervistatore per quanto attiene il tempo impiegato nella codifica.

Si riportano sull'asse delle ordinate i valori della media della variabile chiave prescelta, osservata in diverse situazioni. In questo caso, si riporta la durata media giornaliera ( $\bar{x}_i$ ) della codifica della *Professione* registrata per l'intervistatore sotto osservazione, per ogni giorno di rilevazione disponibile. Sull'asse delle ascisse ci saranno le diverse situazioni osservazionali cui sono riferiti i valori della variabile; in questo caso si tratta dei giorni, ma non è detto che debba esserci sempre una variabile temporale, perché, ad esempio, sull'asse delle ascisse ci potrebbero essere i diversi codificatori e la carta essere riferita ad un solo giorno di rilevazione.

Oltre ai valori osservati si riportano sulla carta anche una retta che rappresenta un valore centrale e due spezzate che rappresentano i limiti di controllo superiore e inferiore. Il valore centrale può essere la media o la mediana dei valori riportati. In questo caso, è la media delle durate medie di codifica dell'intervistatore:

$$\bar{\bar{x}} = \sum \frac{\bar{x}_i n_i}{n_i}$$

dove  $n_i$  è il numero di *Professioni* codificate nel giorno i dall'intervistatore. I limiti di controllo superiore (UCL) e inferiore (LCL) sono quei valori di riferimento che rendono possibile distinguere l'effetto delle cause speciali da quello delle cause comuni. In generale<sup>9</sup>:

---

<sup>9</sup> Viene riportata la formula dei cosiddetti 3sigma limits, dati dal valore centrale  $\pm 3$  volte la deviazione standard, che è quella più comunemente utilizzata. Può anche essere utilizzata una costante moltiplicativa c diversa da 3.

$$UCL = \bar{\bar{x}} + 3 \cdot \hat{S}_{\bar{x}} = \bar{\bar{x}} + 3 \cdot \frac{\hat{S}_x}{n_i}$$

e analogamente

$$LCL = \bar{\bar{x}} - 3 \cdot \hat{S}_{\bar{x}}$$

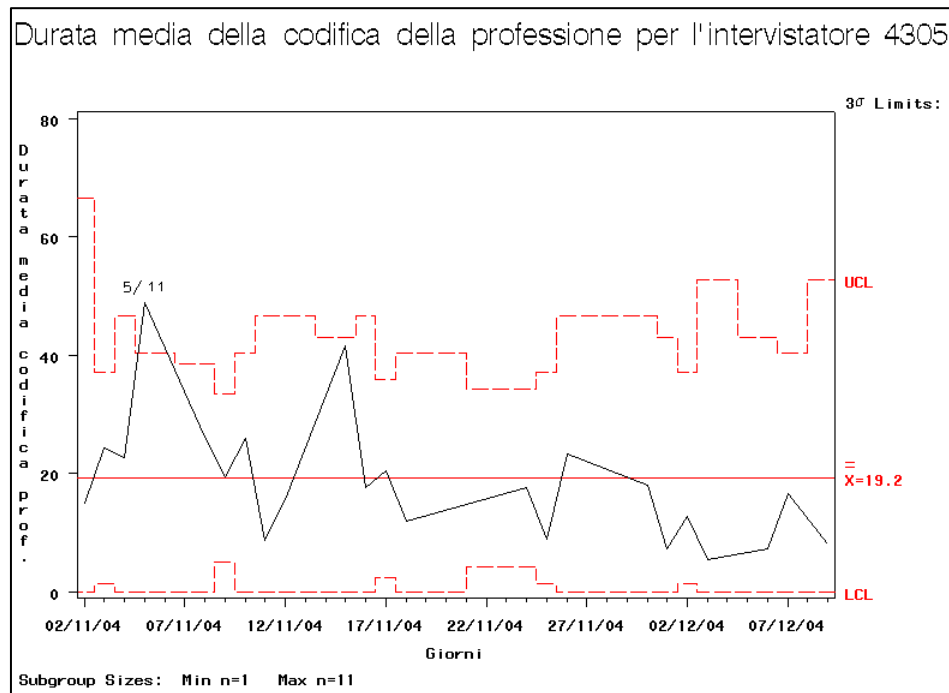
dove  $\hat{S}_{\bar{x}}$  rappresenta la deviazione standard della media giornaliera. La deviazione standard va calcolata per ogni situazione osservazionale tenendo conto del diverso numero di “codifiche” su cui è basata la durata media che è stata calcolata. Poiché il numero di codifiche varia sempre nei diversi giorni (o tra i diversi intervistatori) i limiti di controllo in questa applicazione sono variabili e vengono rappresentati da delle spezzate e non da delle rette parallele all’asse delle ascisse, come avviene quando la media è basata sempre sullo stesso numero di osservazioni.

A questo punto devono essere applicati i test per le cause speciali. Ogni punto al di sopra del limite di controllo superiore o al di sotto del limite di controllo inferiore rappresenta una situazione fuori controllo statistico. La probabilità che si registri un valor medio così distante dalla media generale per solo effetto del caso è talmente bassa da essere necessario un approfondimento sui motivi che l’hanno generato. Altri test per le cause speciali che si applicano solo se sull’asse delle ascisse è rappresentato il tempo sono il test dei *run* (8 o più punti consecutivi tutti al di sopra, o tutti al di sotto del valore centrale) e il test dei *trend* (6 o più punti consecutivi tutti crescenti o tutti decrescenti). Individuate le situazioni fuori controllo statistico bisogna agire per eliminare le cause speciali che le hanno provocate. Per comprendere quale sia la causa che è intervenuta, è possibile che debbano essere effettuate ulteriori analisi e/o approfondimenti.

Una volta che si sono eliminate tutte le possibili fonti di variabilità straordinaria il processo può essere confrontato con le specifiche. In tal caso il valore centrale riportato nella carta può essere un valore di riferimento, cui si vuole che il valor medio della variabile chiave tenda, mentre i limiti, che saranno chiamati limiti di specifica, superiore e inferiore, potranno o essere anch’essi definiti dall’esterno oppure calcolati a partire dai dati, ma considerando il nuovo valore centrale come riferimento. A questo punto il cadere dei punti al di fuori dei limiti indica che la variabilità dovuta a cause comuni è troppo elevata e bisogna agire con un cambiamento radicale nel processo.

Se tutti i punti cadono all’interno dei limiti si è raggiunto un buon risultato, ma il monitoraggio non deve essere interrotto perché le situazioni straordinarie si possono sempre presentare e perché, comunque, devono essere valutati gli effetti delle azioni intraprese per riportare il processo entro i limiti.

Figura 9.1 Carta di controllo di esempio



## 9.2 La procedura di monitoraggio

Le carte di controllo appena descritte rappresentano il fulcro di una intera procedura di monitoraggio della codifica della *Professione* pensata con l'obiettivo di realizzare uno strumento di monitoraggio alternativo alla usuale reportistica.

All'attività di monitoraggio della codifica della *Professione* erano già stati dedicati alcuni dei *report* realizzati per controllare l'andamento della rilevazione. L'esame del contenuto di questi *report* risulta quindi propedeutico alla descrizione della procedura di monitoraggio, al fine di capire le differenze tra i due strumenti di monitoraggio in termini di informazione fornita, semplicità di utilizzo ed efficacia nel cogliere i cambiamenti avvenuti in seguito agli interventi sugli operatori.

I *report* realizzati per le due Indagini sulla *Transizione istruzione-lavoro* sono finalizzati a tenere sotto controllo sia i tempi della codifica sia il ricorso alla codifica a posteriori; in particolare, come indicato nella seguente tabella (Tabella 9), analizzano, per ciascun operatore, il dato giornaliero e cumulato riguardante:

- l'incidenza delle codifiche a posteriori sul totale delle *Professioni* da codificare;
- lo scarto di questa incidenza dal suo valore medio;
- i tempi medi e massimi di codifica;
- i tempi medi e minimi di navigazione nell'applicazione di codifica senza poi attribuire un codice.

Tabella 9- Tempi di codifica della Professione del laureato/diplomato –Cumulato

OPERATORI	Numero di Professioni codificate a posteriori	Numero di Professioni da codificare	Incidenza % di prof codificate posteriori	$\Delta$ rispetto a valore medio	CON CODIFICA Tempo medio	CON CODIFICA Tempo massimo	SENZA CODIFICA Tempo medio	SENZA CODIFICA Tempo minimo
	<b>Totale</b>	<b>592</b>	<b>18190</b>	<b>3.3</b>		<b>0.00.47</b>	<b>0.13.16</b>	<b>0.01.49</b>
1499	10	93	10.8	7.5	0.01.40	0.05.43	0.02.36	0.00.46
1535	1	160	0.6	-2.6	0.00.54	0.07.40	0.03.10	0.03.10
1549	8	169	4.7	1.5	0.00.46	0.05.53	0.02.20	0.01.02
1960		74	0.0	-3.3	0.00.50	0.04.40		
.....								

I *report* contengono quindi le informazioni sufficienti per capire se:

- l'operatore codifica troppo frequentemente a posteriori attribuendo codici non al massimo dettaglio (4 digit);
- i tempi medi di codifica di ciascun operatore sono più o meno lontani dalla media generale;
- il ricorso al modulo di codifica a posteriori è dovuto o ad una scelta autonoma dell'operatore (assolutamente sconsigliata in fase di formazione ed effettuata prima del trascorrere dei due minuti, previsti come tempo massimo) o di non allungare i tempi di intervista, codificando al termine di essa (bassi tempi di codifica senza attribuire un codice) o alla sua incapacità di codificare *on-line* (tempi alti di codifica senza attribuire un codice).

Incidenze troppe alte di codifiche a posteriori e tempi troppo lunghi senza poi attribuire un codice suggeriscono di monitorare l'operatore presso la ditta e di effettuare un supplemento di spiegazione sulle modalità di codifica o di invitare l'operatore ad utilizzare la codifica a posteriori non come procedura standardizzata, perché ritenuta da lui più comoda ai fini della conclusione dell'intervista, ma solo nei casi di effettiva necessità.

Sebbene i *report* appena descritti siano un ottimo strumento di monitoraggio, presentano alcune difficoltà di lettura e alcune lacune informative, infatti:

- non sono di facile analisi perché riportano tante righe quanti sono gli operatori presenti nell'Indagine, anche quelli che hanno lavorato un solo giorno;
- non indicano quando un operatore si discosti significativamente dalla media;
- non indicano se il ricorso al modulo di codifica a posteriori sia dovuto a codifiche non complete o a mancate codifiche;
- riportano l'informazione cumulata e giornaliera, ma non l'andamento nel tempo. Diventa quindi difficile capire, se, una volta individuato il problema dell'operatore ed effettuato un *de-briefing*, questi abbia effettivamente recepito le spiegazioni. Il dato cumulato, infatti, in quanto tale, non riesce ad evidenziare immediatamente i cambiamenti di tendenza che invece emergerebbero subito da un grafico di *trend*;
- non chiarificano se tempi troppo bassi per codifiche al massimo dettaglio siano il frutto della bravura dell'operatore o la conseguenza di attribuzioni di codici simili o afferenti sempre allo stesso Grande Gruppo sul quale si verificano concentrazioni elevate rispetto all'andamento medio.

Nel disegnare la procedura di monitoraggio, si è tenuto conto del fatto che questa non dovesse dare adito giornalmente ad analisi eccessivamente approfondite su ciascun operatore, perché non lascerebbero il tempo per l'applicazione dei loro risultati, ossia per gli interventi da effettuare sul campo. Si è voluto, quindi, sperimentare una procedura che, data la complessità del processo di codifica assistita della *Professione*, dato l'alto numero di operatori telefonici da monitorare e data la necessità di utilizzare nel modo più efficiente le potenzialità offerte dalle Carte di Controllo fosse:

- di immediata e di facile lettura;
- sintetica ed efficace, ossia riuscisse ad analizzare in modo sintetico più indicatori connessi con la codifica della *Professione*;
- totalmente strutturata, ossia definisse a priori e in modo oggettivo 'quando' un operatore è da controllare e 'come' correggere l'operatore, ossia come riportare il suo processo di codifica nei limiti di controllo dai quali è uscito.

La procedura di monitoraggio realizzata consiste nell'identificare:

- a) le variabili chiave del processo di codifica;
- b) le regole per la definizione del fuori controllo e le azioni da espletare per riportare il processo nei limiti;
- c) eventuali analisi *ad hoc* per approfondire le motivazioni di determinate tipologie di fuori controllo.

Le variabili chiave individuate per monitorare la codifica della *Professione* sono:

- 1) la durata della codifica;
- 2) la qualità della codifica;
- 3) il tasso di risposta.

La durata della codifica è stata definita come il tempo che intercorre dal momento in cui si pone il quesito sulla *Professione* fino all'attribuzione del codice (o dell'invio nel caso di codice non assegnato) nel campo predisposto. Si tratta dello stesso indicatore analizzato nei report.

La qualità della codifica è stata definita come il rapporto tra il totale delle codifiche al massimo dettaglio (4 digit) e il totale delle *Professioni* da codificare.

Il tasso di risposta è stato definito come il rapporto tra le interviste complete e la somma delle interviste complete, rifiuti e interruzioni definitive.

Per ciascuna variabile chiave è stata costruita la relativa carta di controllo; queste carte, dette di *screening*, vengono elaborate settimanalmente e in caso di operatore fuori controllo per una o più delle variabili chiave, questo viene monitorato in modo più dettagliato, ossia viene osservato l'andamento dell'operatore nel tempo. Il tipo di approfondimento per l'operatore dipende dal tipo di fuori controllo; gli approfondimenti per operatore, previsti dalla procedura di monitoraggio, sono per il:

- fuori controllo per "Durata Codifica Professione Alta";
- fuori controllo per "Durata Codifica Professione Bassa";
- fuori controllo per "Qualità Codifica Professione Bassa";
- fuori controllo per "Tasso di risposta Basso"



Per ogni approfondimento, la procedura prevede a monte la tipologia di analisi da effettuare per capire la motivazione del fuori controllo e cosa fare per riportare il processo nei limiti, come descritto nei paragrafi seguenti.

### 9.2.1 *Approfondimento per operatore fuori controllo per 'Durata Codifica Professione'*

L'approfondimento in caso di fuori controllo per un tempo di codifica troppo alto o troppo basso, prevede che l'operatore debba essere monitorato analizzando la 'Qualità della codifica'. Dalle combinazioni dei risultati su questi due indicatori (Figura 9.2), seguono interventi diversi; in particolare:

- Se la durata e la qualità sono alte occorre verificare, prima di tutto, se ciò non abbia comportato un incremento nelle interruzioni definitive dopo il quesito sulla *Professione*<sup>10</sup>.

- se ciò si verifica, l'operatore deve essere addestrato di nuovo sulla codifica, focalizzando il *de-briefing* sull'utilizzo di quelle particolari accortezze (ad esempio l'uso delle parole chiave, *cf* paragrafo 6) che, se rispettate, consentono di individuare più velocemente il codice da attribuire mantenendo lo stesso standard di qualità.

- qualora un tempo di codifica alto non incida sulla percentuale di interruzioni definitive, non occorre intervenire sull'operatore, in quanto ha semplicemente bisogno di un tempo più elevato della media per raggiungere il massimo della qualità di codifica (codici al massimo dettaglio).

- Se la durata è alta e la qualità è bassa, l'operatore deve essere addestrato di nuovo sulla *Professione*, sia sui criteri della Classificazione che sull'utilizzo dello strumento di codifica, in quanto impiega molto tempo sul quesito della *Professione* per poi attribuire codici generici (a meno di quattro digit). In questa situazione di fuori controllo, un'enfasi maggiore alla necessità di un intervento sull'operatore viene data dalla presenza di non codifiche: questo viene analizzato attraverso un istogramma che riporta la distribuzione dei casi classificati nei nove Grandi Gruppi e in quello delle 'non codifiche' per il singolo operatore e per il totale.

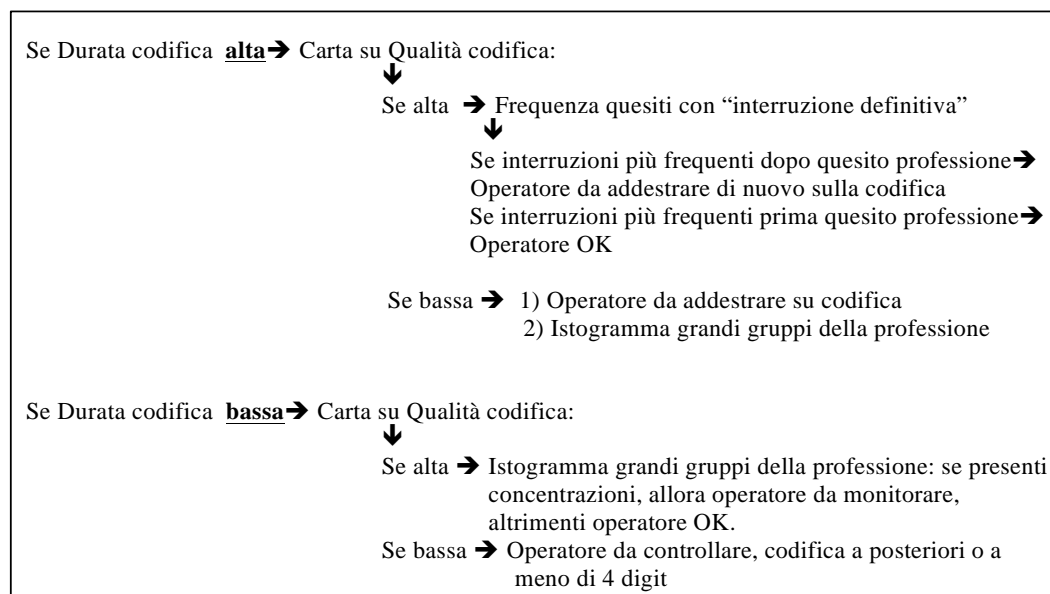
- Se la durata della codifica è bassa e la qualità è alta, occorre controllare che questa situazione, in linea di principio ideale, sia il frutto della bravura dell'operatore, e non la conseguenza di codifiche simili tra loro e quindi concentrate in particolari Grandi Gruppi. Questa analisi si effettua attraverso l'istogramma sopra citato: se non emergono concentrazioni, non dovrà essere effettuato alcun intervento, mentre, in caso contrario, se le concentrazioni supereranno di una percentuale stabilita la media generale, allora l'operatore dovrà essere monitorato in corso di rilevazione (previa un'analisi dei testi inseriti e dei codici assegnati) e sollecitato ad abbandonare questo tipo di codifica.

- Se durata e qualità della codifica sono basse, l'operatore dovrà essere monitorato in corso di rilevazione perché cerca di accelerare i tempi di intervista a scapito della qualità dell'informazione rilevata.

---

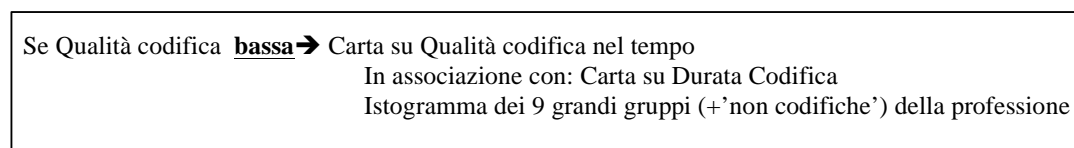
<sup>10</sup> L'uso della codifica *on-line* per variabili con Classificazioni complesse come la Professione è, infatti, sconsigliato quando possa incidere negativamente sui tempi di intervista incrementando la percentuale di interruzioni definitive e contraendo conseguentemente il tasso di risposta (Macchia, Murgia 2002).

Figura 9.2 Schema di approfondimento per 'Durata Codifica Professione'



### 9.2.2 Approfondimento per operatore in caso di fuori controllo per 'Qualità Codifica Professione'

Figura 9.3 Schema di approfondimento per 'Qualità Codifica Professione'



In caso di fuori controllo per qualità bassa della codifica, la procedura suggerisce di effettuare l'analisi temporale dell'indicatore a prescindere dall'andamento delle altre variabili, al fine di verificare se il fuori controllo sia un evento episodico o che si ripete nel tempo. In quest'ultimo caso, ossia quando una qualità bassa della codifica rappresenta un evento sistematico, l'analisi congiunta con la "Durata della codifica" e dell'istogramma permetterà di diversificare gli interventi sull'operatore in modo analogo a quanto descritto nell'approfondimento precedente.

In corso d'opera, per l'Indagine sui diplomati, è stata sollevata la necessità di completare l'indicatore sulla 'Qualità della codifica' (codifiche al massimo dettaglio sul totale delle codifiche) con quello sulla coerenza tra 'Posizione nella Professione' e 'Grande Gruppo della Professione' (ossia primo digit del codice assegnato). Il nuovo indicatore sulla qualità verrà messo a punto per la prossima edizione della procedura e sarà probabilmente monitorato sulla base di un altro strumento di controllo quale il 'diagramma di Pareto' (WESTAT. 1992).

### 9.2.3 Approfondimento per operatore fuori controllo per 'Tasso di Risposta Basso'

Il caso di fuori controllo sul 'Tasso di risposta' rappresenta la situazione più complessa che esamina, oltre alle variabili strettamente legate alla Professione, anche quelle connesse con gli esiti di contatto. Questa parte della procedura monitora l'eventualità che tempi di codifica troppo lunghi comportino dei bassi tassi di risposta: in questo

caso l'analisi di approfondimento (Figura 9.4) suggerisce di controllare la durata dell'intervista per capire se la lunghezza eccessiva sia la causa di questo risultato:

- Se la durata dell'intervista è alta, viene esaminata la durata della codifica per capire se sia questa la componente che pesa maggiormente sui tempi di rilevazione. L'esame di questo indicatore viene effettuato congiuntamente a quello della qualità: oltre alle diverse azioni da effettuarsi sul campo che derivano dai modi in cui possono combinarsi i due indicatori, già descritte nel caso di fuori controllo per la "Durata della codifica", la procedura suggerisce di:

- addestrare di nuovo l'operatore quando il tasso di risposta basso è legato a tempi di codifica alti e ad una qualità bassa;

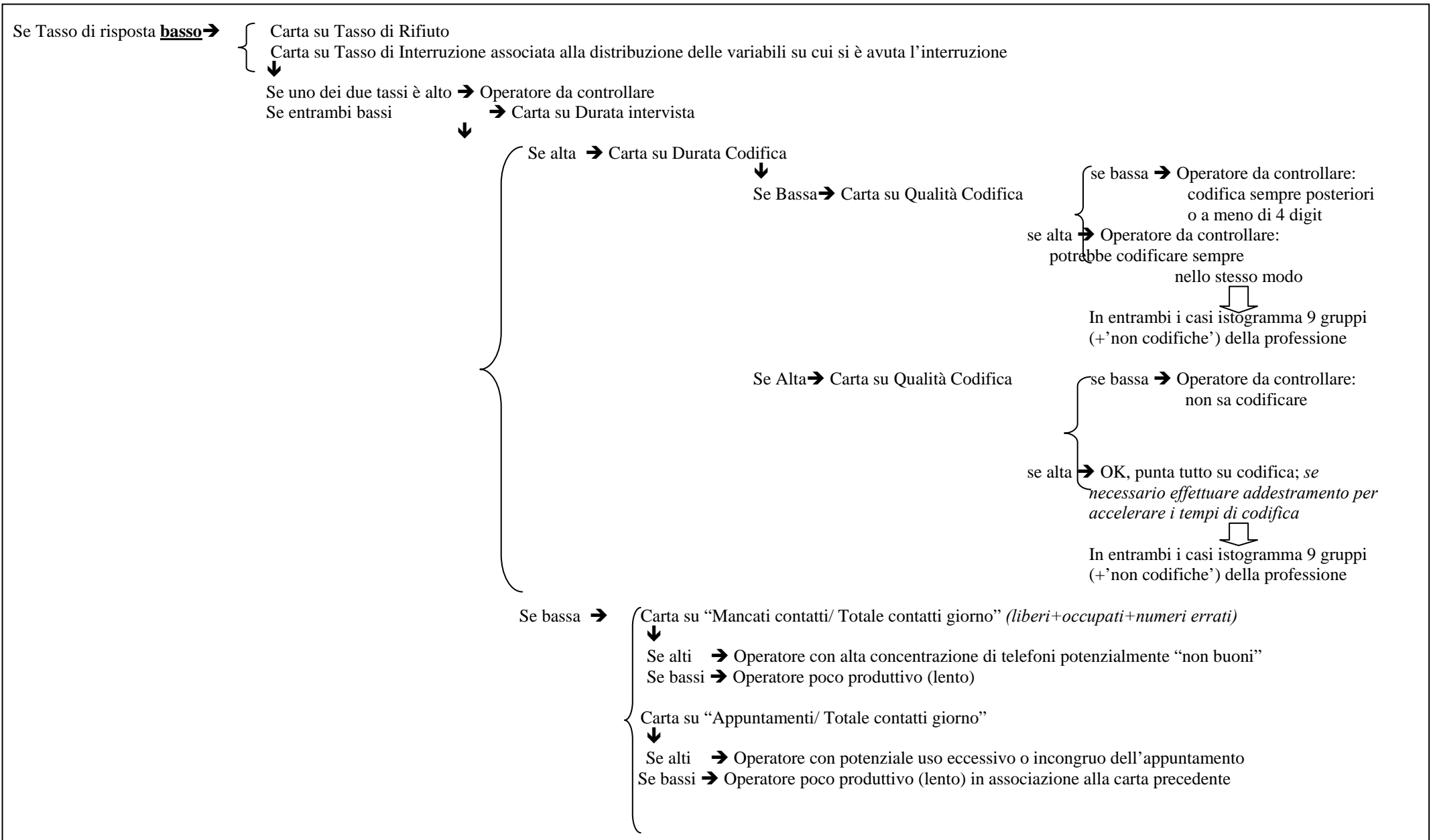
- di osservare l'operatore sul campo al fine di individuare e correggere quegli atteggiamenti che, pur garantendo livelli di qualità ottimali, allungano eccessivamente i tempi di codifica.

- Se la durata dell'intervista è bassa, le cause di un tasso di risposta basso non sono più da ricercarsi nella codifica della *Professione*, ma nella gestione dei contatti telefonici. In questo caso, la procedura suggerisce di analizzare le carte sugli indicatori dei "Mancati contatti" e degli "Appuntamenti":

- se uno degli indicatori è alto, allora l'operatore è da controllare sul campo perché potrebbe fare un uso incongruo di questi esiti di contatto (libero, occupato, numero errato, appuntamento) ;

- se entrambi gli indicatori sono bassi, allora è semplicemente la scarsa produttività dell'operatore ad influire negativamente sui livelli del tasso di risposta.

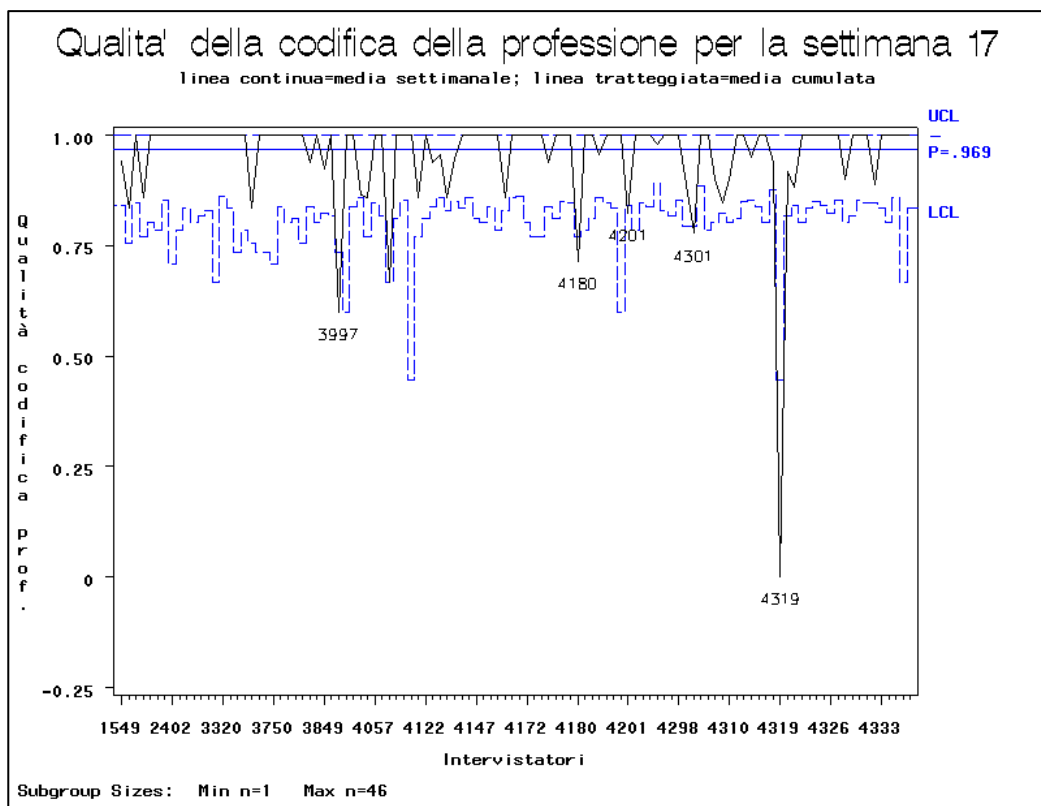
Figura 9.4 Schema di approfondimento per “Tasso di risposta basso”



#### 9.2.4 La procedura di monitoraggio in termini applicativi

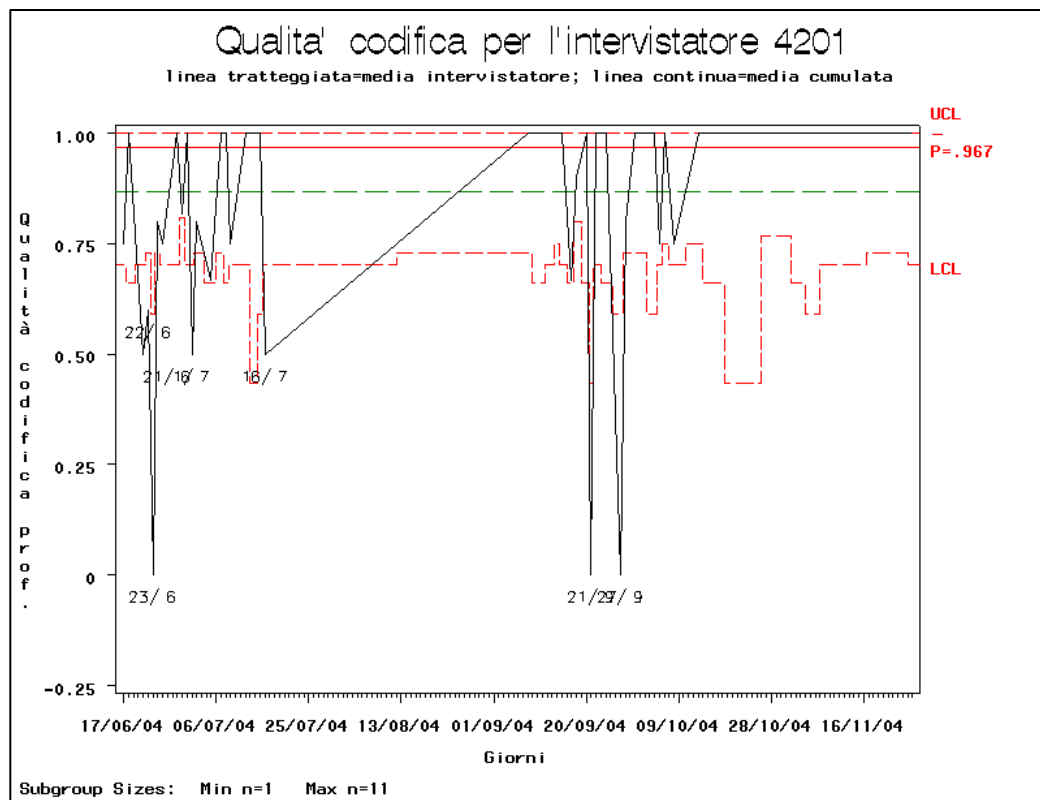
Da quanto descritto, emerge quindi, che oltre alle tre carte di controllo relative alle variabili chiave, vengono prodotte altre carte che servono ad approfondire le situazioni di fuori controllo sulle prime. Le carte costruite per ciascuna variabile chiave, che, come già detto, sono identificate con il nome di carte di *screening*, forniscono l'andamento giornaliero e/o settimanale dei tre indicatori per tutti gli operatori attivi nel periodo selezionato (giorno o settimana). Per queste carte i limiti di controllo sono calcolati rispetto alla media giornaliera o settimanale, mentre la media generale viene solo riportata sul grafico come ulteriore informazione.

Figura 9.5 Carta di screening per "Qualità della codifica" nella settimana 17°



Qualora un operatore risulti fuori dai limiti per uno o più dei tre indicatori di *screening*, viene monitorato nel tempo attraverso le carte di approfondimento che mostrano sempre l'andamento giornaliero dell'indicatore (dal primo all'ultimo giorno di lavoro dell'operatore) a prescindere dal tipo di *screening* scelto (giornaliero o settimanale). Per le carte di approfondimento i limiti sono calcolati prendendo come riferimento la media generale e non la media del singolo operatore, che viene comunque rappresentata sul grafico come supporto interpretativo.

Figura 9.6 Carta di approfondimento per operatore 4201 per “Qualità della Codifica”



In queste carte, quindi, il processo di codifica del singolo operatore viene confrontato con il processo di codifica medio generale (valore desiderato) che è calcolato su tutti gli operatori, dall’inizio della rilevazione fino al giorno di *screening*; la scelta della media generale come valore desiderato e non della media del singolo operatore è insita nel fatto che molto spesso l’operatore registra un processo “stabile”, ossia contenuto sempre all’interno dei propri limiti, ma costantemente fuori dai limiti di controllo desiderati.

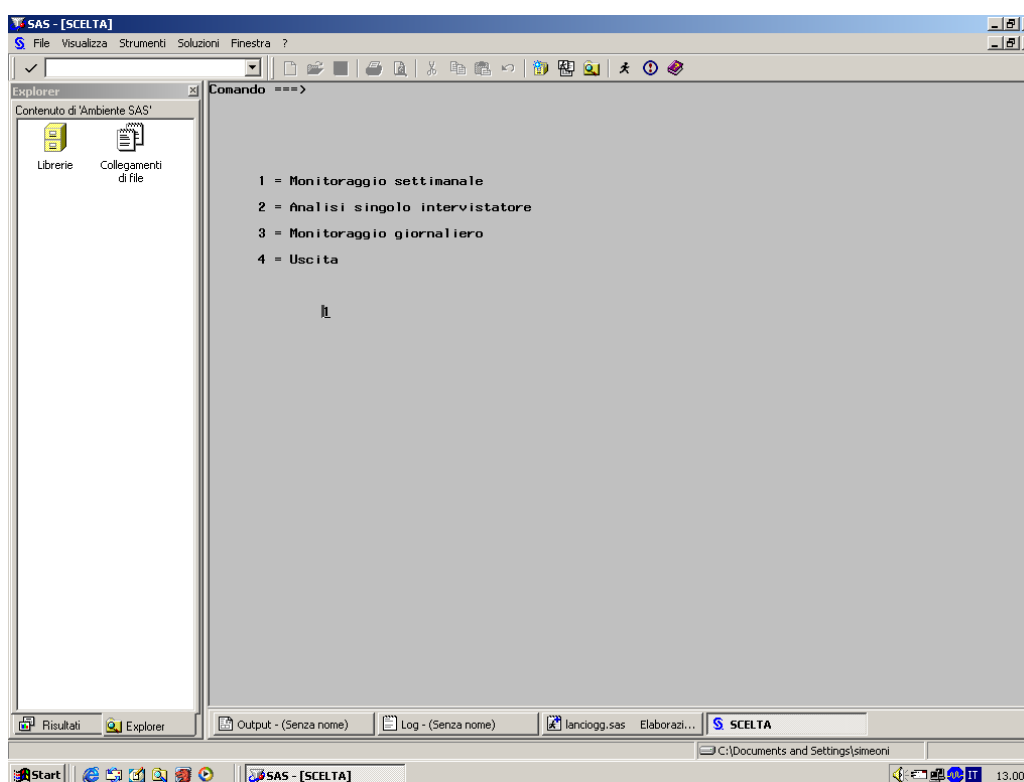
Attraverso questa sequenza operativa è possibile, quindi, monitorare sia le nuove situazioni di fuori controllo che periodicamente vengono evidenziate dalle carte di *screening*, sia analizzare l’andamento nel tempo dell’operatore: questo, oltre ad offrire un quadro di analisi più completo rispetto ad un dato cumulato, permette di verificare se le azioni effettuate sul campo, quali *de-briefing*, controlli, ecc., abbiano avuto gli effetti sperati riportando l’operatore all’interno dei limiti desiderati.

L’intera procedura è stata sviluppata con il modulo QC del software SAS. Un esempio dell’applicazione è riportato nel paragrafo successivo.

### 9.3 La procedura in SAS

In figura 9.7 è riportata la schermata iniziale dell'applicazione, che consente di scegliere quale funzione si vuole attivare. La strategia di monitoraggio sviluppata prevede di effettuare il monitoraggio settimanale (opzione 1), ma l'applicazione consente anche di passare direttamente ad analizzare il comportamento di un singolo intervistatore (opzione 2) oppure di costruire le carte di *screening* con riferimento ad un giorno e non ad una settimana (opzione 3). Analizzare direttamente il comportamento di un intervistatore può essere utile per verificare a posteriori l'esito di interventi effettuati a seguito di monitoraggi precedenti, mentre le carte di *screening* giornaliere possono essere utilizzate per un monitoraggio più puntuale.

Figura 9.7 Schermata iniziale dell'applicazione



Selezionando l'opzione 1 si passa alla schermata di scelta della settimana di rilevazione da monitorare, riportata in figura 9.8. E' sufficiente inserire il numero della settimana di rilevazione (in questo esempio la 17, che corrisponde alla settimana dal 27 settembre al 2 ottobre 2004) e l'applicazione costruisce le carte di *screening* relative alle variabili chiave.

A titolo esemplificativo, si riporta in figura 9.9 la carta di *screening* relativa alla variabile chiave 'Qualità della codifica'. La carta evidenzia cinque operatori fuori controllo statistico di cui vengono riportati i codici identificativi che serviranno in seguito a richiedere la costruzione delle carte di approfondimento.

Figura 9.8 Schermata di scelta della settimana di rilevazione da monitorare

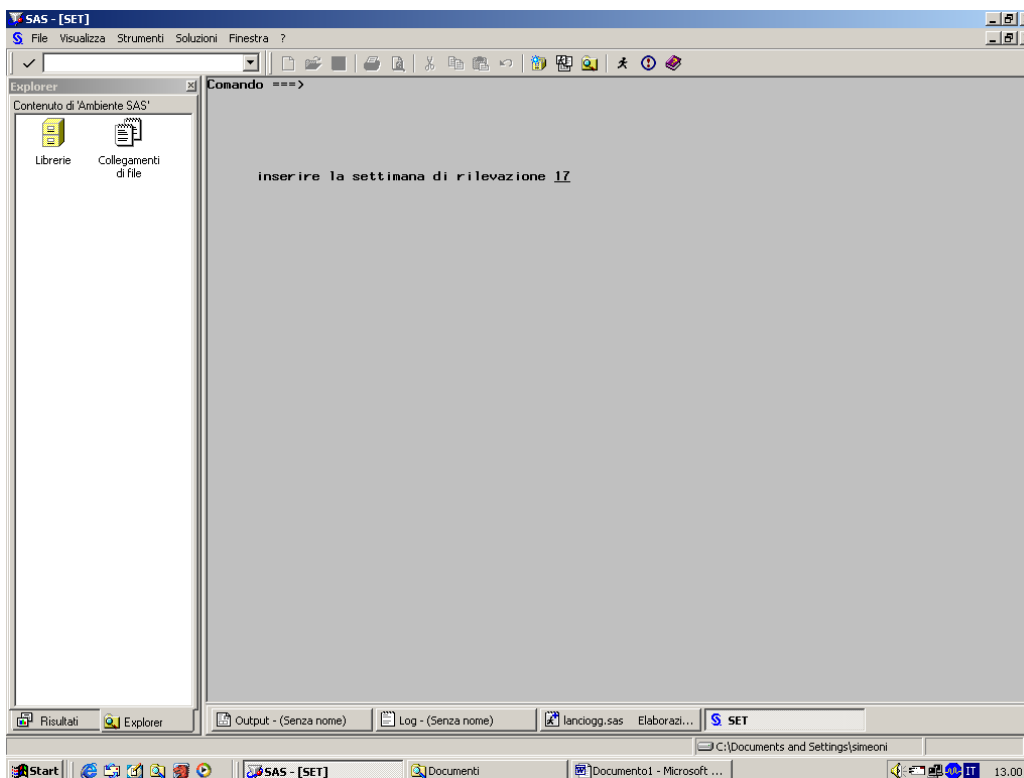
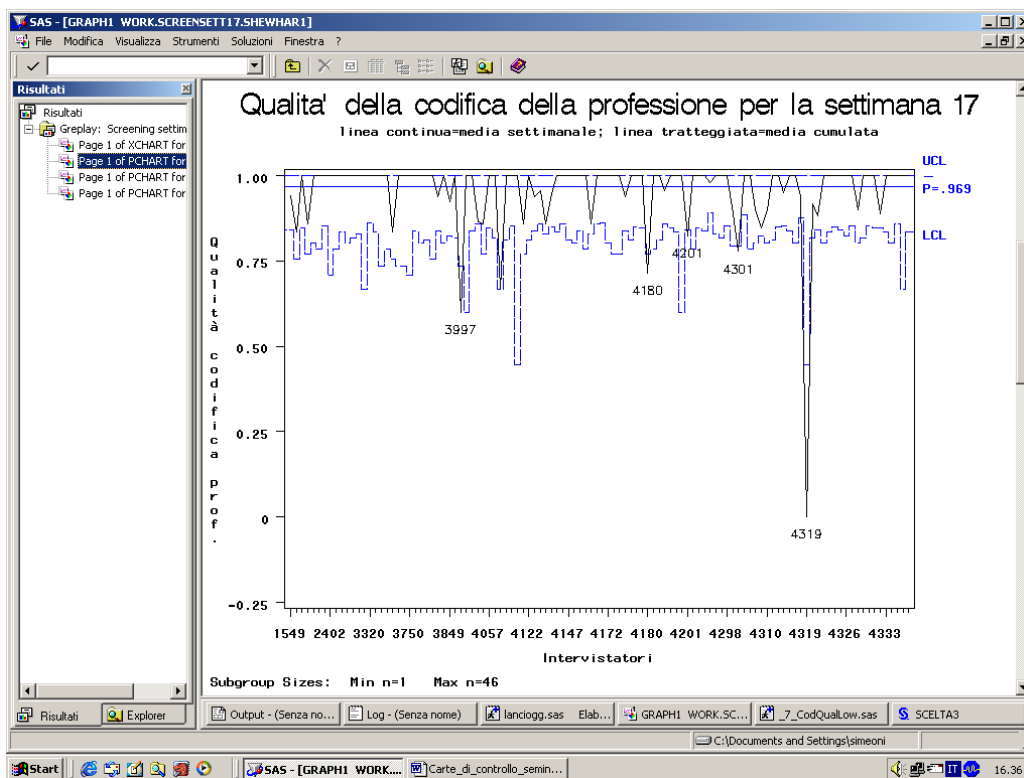


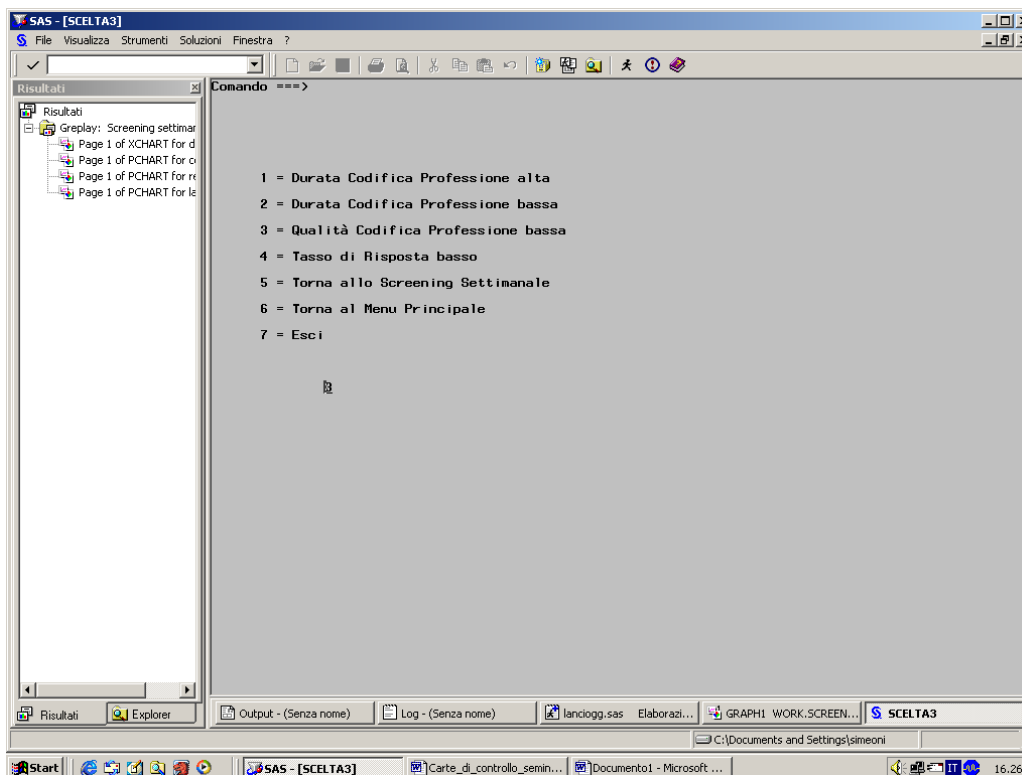
Figura 9.9 Schermata della carta di screening per "Qualità della codifica" nella settimana richiesta





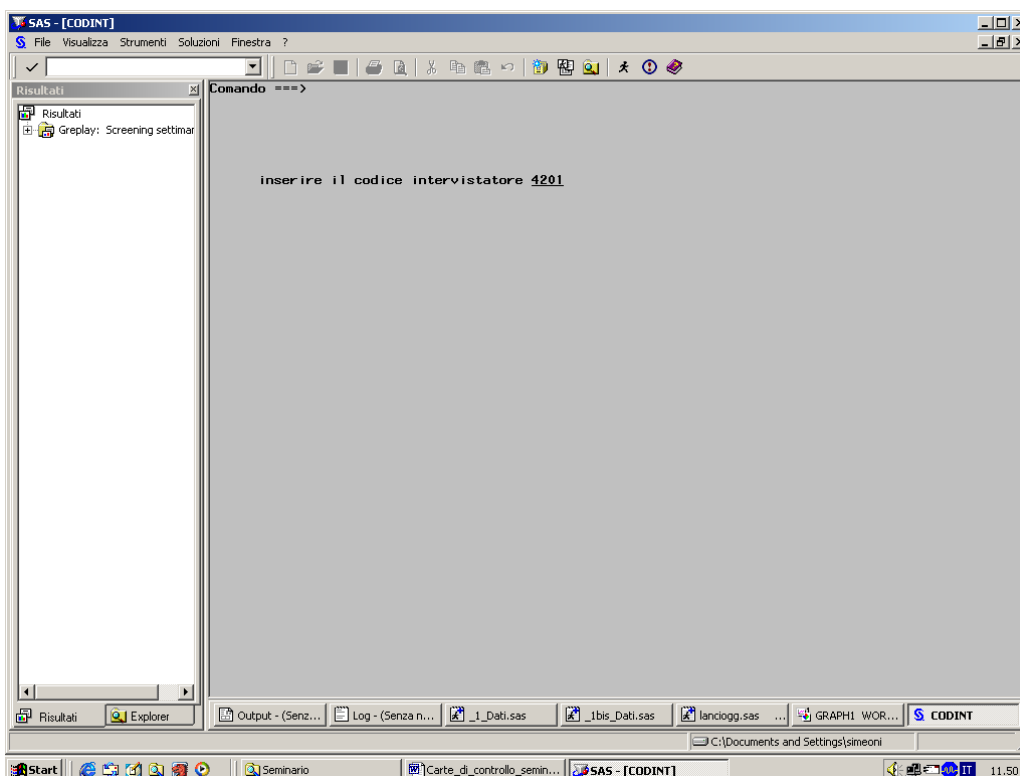
Una volta individuati i codificatori fuori controllo statistico, è possibile passare alla schermata riportata in figura 9.10 in cui si può richiedere l'approfondimento più appropriato: se l'intervistatore è risultato fuori dai limiti di controllo per 'Durata della codifica' troppo elevata, si sceglierà l'opzione 1, se per 'Durata della codifica' troppo bassa, si sceglierà l'opzione 2 etc.

Figura 9.10 Schermata di scelta del tipo di approfondimento



Selezionato il tipo di approfondimento desiderato, la procedura richiede di immettere il codice dell'operatore da monitorare (figura 9.11). Nell'esempio riportato è stato richiesto l'approfondimento per Qualità della codifica bassa per l'intervistatore 4201.

Figura 9.11 Schermata di scelta dell'intervistatore da monitorare



Nelle figure seguenti sono riportati i grafici previsti per l'approfondimento nel caso di fuori controllo per 'Qualità della codifica bassa'. In particolare, in figura 9.12 è rappresentata la carta relativa all'andamento nel tempo della variabile 'Qualità della codifica'; in figura 9.13 l'andamento nel tempo della variabile 'Durata della codifica' e in figura 9.14 sono, invece, riportati gli istogrammi sovrapposti che rappresentano la distribuzione dei Grandi Gruppi della *Professione* per il totale degli operatori e per l'operatore in esame (la modalità 0 corrisponde alle mancate codifiche *on line*).

Si può notare come l'operatore 4201 presentasse bassi livelli di qualità con moltissimi giorni fuori controllo fino all'inizio del mese di ottobre 2004; sempre in questo periodo, presentava dei tempi di codifica superiori alla media, anche se non sempre fuori controllo statistico. L'istogramma mostrava, poi, una tendenza spiccata a codificare a posteriori. Da tale analisi (effettuata nella settimana a cavallo tra la fine di settembre e l'inizio di ottobre) risultava piuttosto chiaro che l'operatore trovava molta difficoltà nel codificare, pur investendo tempo nel tentativo di farlo. Di conseguenza, il 5 ottobre 2004, è stato effettuato un intervento formativo *ad hoc* sulla codifica della *Professione*. Le carte evidenziano l'efficacia di tale intervento: la qualità della codifica è diventata ottima e il tempo di codifica si è ridotto rientrando sempre nei limiti di controllo.

Figura 9.12 Carta di approfondimento per "Qualità della codifica" per l'intervistatore richiesto

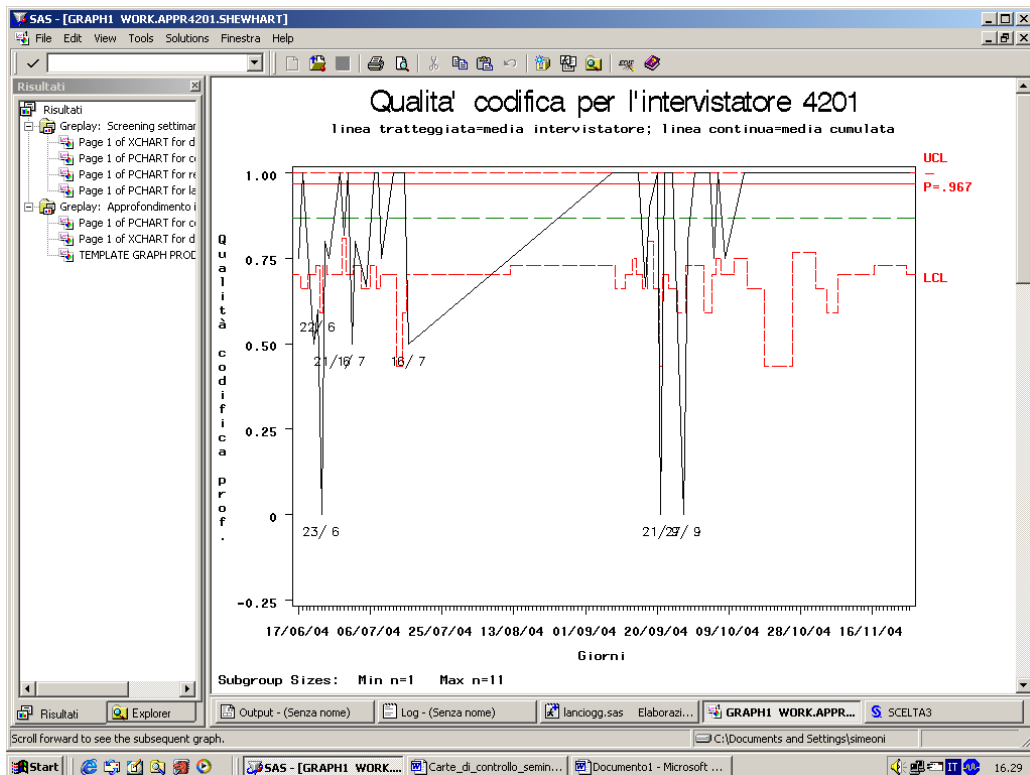


Figura 9.13 Carta di approfondimento per "Durata della codifica" per l'intervistatore richiesto

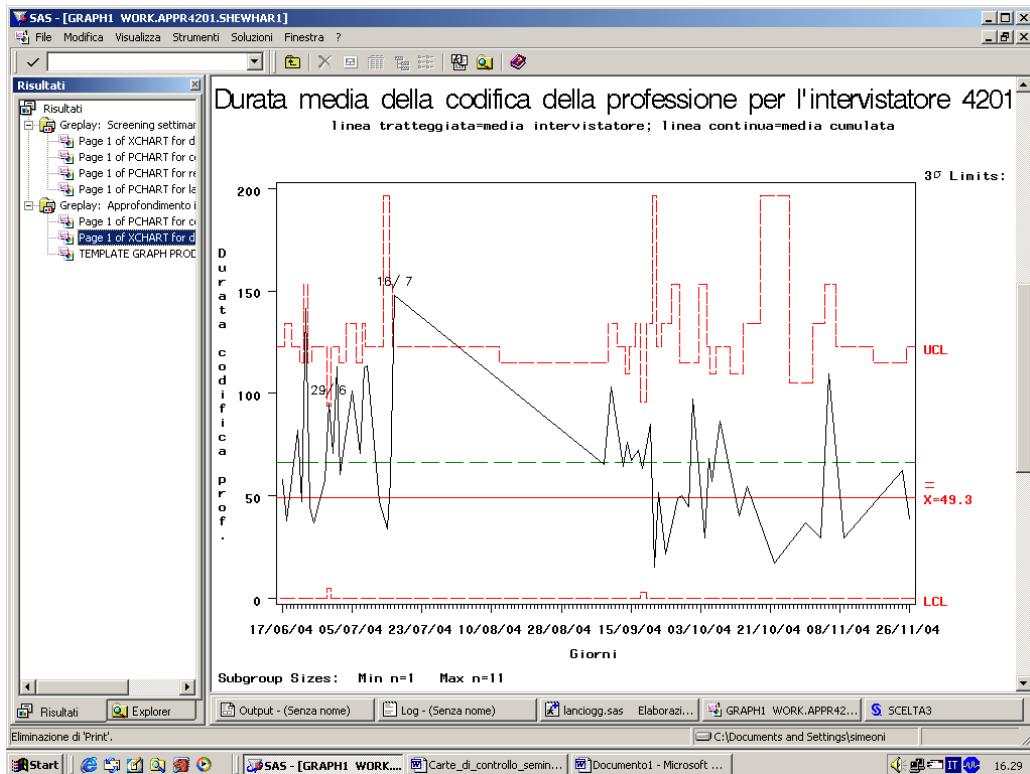
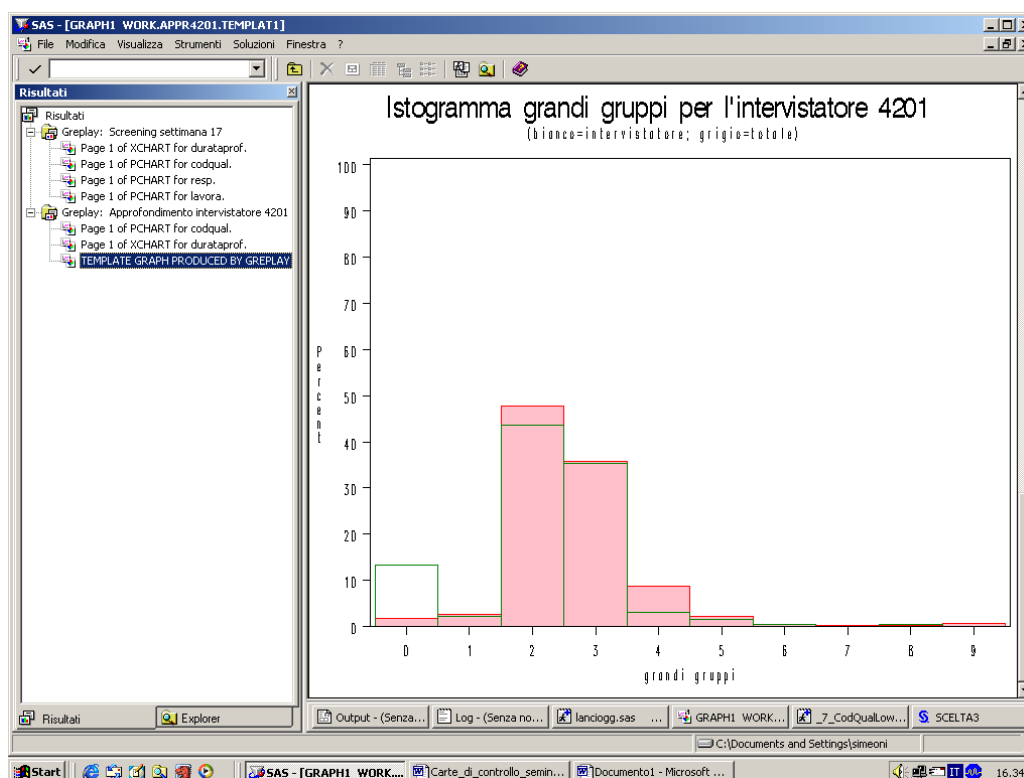


Figura 9.14 Istogramma Grandi Gruppi della Professione per l'intervistatore richiesto



Una volta effettuata l'analisi dell'approfondimento di un operatore, è possibile tornare alla schermata riportata in figura 9.10 per proseguire o con un altro approfondimento per lo stesso o per un altro operatore, oppure tornare alla fase di *screening* o infine uscire dall'applicazione.

Concludendo sulle caratteristiche dell'applicazione, si evidenzia come essa gestisca gli output prodotti in maniera razionale: nella parte sinistra della figura 9.14 è possibile osservare come i grafici vengano automaticamente inseriti in cartelle con nomi auto-esplicativi, quali "Screening settimana xx" oppure "Approfondimento intervistatore xxxx". Questo risulta molto utile quando vengono prodotti approfondimenti per diversi operatori e di conseguenza il numero dei grafici costruiti diventa molto elevato.

#### 9.4 Conclusioni

Le carte di controllo permettono, a nostro avviso, di avere a disposizione uno strumento di monitoraggio molto efficace in quanto, pur analizzando contemporaneamente un insieme composto di variabili, sono molto semplici nell'interpretazione e nella verifica dell'efficacia degli interventi effettuati sul campo, grazie soprattutto al supporto interpretativo offerto dalla grafica. Migliorano la reportistica, da cui prendono spunto, in quanto:

- permettono di individuare immediatamente gli operatori che sono probabilmente da controllare;

- mostrano l'andamento nel tempo degli indicatori;
- mostrano la presenza di eventuali concentrazioni di codifiche attraverso l'istogramma;
- indicano l'incidenza non solo delle codifiche parziali, ma anche delle non codifiche;

Infine, esaminando gli aspetti positivi e negativi connessi all'utilizzo di questo strumento è possibile dire che i principali vantaggi sono:

- l'immediatezza nell'identificare le situazioni di fuori controllo;
- la possibilità di correggere il processo in corso d'opera e di verificare in tempi brevissimi l'efficacia delle azioni migliorative espletate sul campo;
- l'adattabilità dell'applicativo a tutte le Indagini CATI;
- l'utilizzo del software SAS diffuso in ISTAT.

Al contempo, i principali limiti legati al loro utilizzo sono:

- il costo annuale della licenza del modulo QC,
- l'applicabilità ad Indagini CATI che coprono un arco temporale di almeno due mesi di rilevazione, ossia di un tempo necessario a cogliere, correggere e valutare le situazioni di fuori controllo.

## 10 L'esperienza dell'adozione del modulo di codifica assistita: le Indagini sulla *Transizione istruzione-lavoro*

### 10.1 I principali risultati delle Indagini 2004 sulla *Transizione istruzione-lavoro*

Le ultime Indagini condotte dall'ISTAT sulla *Transizione istruzione-lavoro*, lo ricordiamo, hanno avuto come oggetto i laureati e i diplomati di scuola secondaria superiore dell'anno 2001 intervistati a circa tre anni dal conseguimento del titolo.

Per l'indagine sull'*Inserimento professionale dei laureati* sono state realizzate circa 26mila interviste complete, con un tasso di risposta complessivo pari al 67,5%, superiore di circa 10 punti percentuali rispetto a quello dell'indagine 2001<sup>11</sup>. Il tasso di rifiuto<sup>12</sup> è sceso dal 4,0% al 3,6% e anche le interruzioni<sup>13</sup> sono diminuite, rappresentando solo lo 0,6% delle contatti telefonici (cfr. tabella 10.1).

Tabella 10.1 – Indagine sull'inserimento professionale dei laureati - Principali risultati

	<i>Indagine 2004</i>	<i>Indagine 2001</i>
<i>Campione</i>	37.778	36.376
<i>Interviste complete</i>	25.510	20.844
<i>Tasso di risposta</i>	67,5%	57,3%
<i>Tasso di rifiuto</i>	3,6%	4,0%
<i>Tasso di interruzione</i>	0,6%	1,4%

Nell'*Indagine sui percorsi di studio e di lavoro dei diplomati* si sono realizzate circa 20.400 interviste con un tasso di risposta che ha fatto registrare un incremento di quasi 20 punti percentuali rispetto all'edizione precedente (70,5% contro 51,7%); il tasso di rifiuto si è più che dimezzato, scendendo dal 10,8% al 4,8%, mentre il tasso di interruzione è rimasto su valori pressoché trascurabili (cfr. tabella 10.2).

<sup>11</sup> In totale sono state effettuate 26.006 interviste, di queste circa 500 riguardano laureati nei nuovi corsi di laurea triennali che non vengono considerati nell'analisi.

<sup>12</sup> Il tasso di rifiuto è dato dal rapporto tra rifiuti e la somma di complete, rifiuti e interruzioni.

<sup>13</sup> Il tasso di interruzione è dato dal rapporto tra interruzioni e la somma di complete, rifiuti e interruzioni.

Tabella 10.2 – Indagine sui percorsi di studio e di lavoro dei diplomati - Principali risultati

	<i>Indagine 2004</i>	<i>Indagine 2001</i>
<i>Campione</i>	29.073	45.033
<i>Interviste complete</i>	20.408	23.262
<i>Tasso di risposta</i>	70,5%	51,7%
<i>Tasso di rifiuto</i>	4,8%	10,8%
<i>Tasso di interruzione</i>	0,5%	0,7%

Il forte incremento del tasso di risposta registrato in entrambe le Indagini si deve attribuire principalmente al miglioramento delle liste di base dalle quali sono stati estratti i campioni di giovani da intervistare. La più diffusa informatizzazione degli archivi delle scuole e degli atenei ha permesso, infatti, la costruzione di liste più accurate che hanno facilitato il reperimento delle unità campionarie. In particolare, per l'Indagine dei diplomati è stata anche introdotta una nuova modalità di raccolta dei nominativi presso le scuole che consisteva nell'inserimento dei dati via web (il 40% dei nominativi è stato fornito in questo modo). Un ulteriore elemento che ha senza dubbio contribuito ad innalzare il tasso di risposta, in particolare nell'Indagine sui diplomati, è stata una più efficiente gestione degli appuntamenti da parte del sistema di gestione dei contatti telefonici (BLAISE), che ha permesso di effettuare più di un quarto delle interviste in fasce orarie diverse da quella consueta serale (26,4% contro il 9,1% dell'indagine 2001). A questo poi si deve aggiungere l'effetto della diffusione sempre più frequente dei numeri di telefono cellulare, in modo particolare tra i giovani, usati a volte anche in sostituzione del telefono fisso, che ha facilitato il contatto telefonico ed ha permesso di rintracciare numerosi rispondenti che altrimenti sarebbe stato difficile contattare. Questo fenomeno è apparso evidente soprattutto nell'Indagine sui laureati, dove circa il 28% delle interviste complete è stata realizzata grazie ad un contatto tramite telefono cellulare (nell'Indagine sui diplomati la percentuale è stata di circa il 13%).

Questi primi dati confermano dunque i risultati attesi per entrambe le rilevazioni. Nonostante l'inevitabile incremento della durata media delle interviste dovuto in buona parte all'introduzione della codifica assistita della *Professione* (nell'Indagine sui laureati la durata media delle interviste è passata dai 10 minuti e 56 secondi a 12 minuti e 46 secondi, in quella sui diplomati invece, l'aumento è stato più contenuto, da 13 minuti e 20 secondi a 13 minuti e 43 secondi) (cfr. tabella 10.3) non solo la partecipazione alle due Indagini è stata maggiore rispetto al 2001, ma le interviste che si sono interrotte nel corso dei quesiti sulla *Professione* (dell'intervistato o del padre) sono risultate pari a poche unità.

Tabella 10.3 – Indagini sulla Transizione istruzione-lavoro – Durata media delle interviste

	<i>Indagine 2004</i>	<i>Indagine 2001</i>
<i>Indagine Laureati</i>	12' 46"	10' 56"
<i>Indagine Diplomati</i>	13' 43"	13' 20"

## 10.2 Il monitoraggio della codifica della Professione

Gli indicatori predisposti per il monitoraggio della codifica della *Professione* durante la fase di rilevazione, possono riassumersi nei seguenti punti:

- numero di codifiche a posteriori e loro incidenza sul totale delle *Professioni* rilevate;
- dettaglio raggiunto nelle codifiche (ossia il numero di digit);
- distribuzione delle codifiche effettuate per Grande Gruppo;
- tempi medi e massimi di codifica della *Professione* in corso di intervista;
- tempi medi e minimi di permanenza nel modulo di codifica assistita nei casi di mancata codifica in corso di intervista.

L'analisi quotidiana di questi indicatori, riferiti in particolar modo alla *Professione* dell'intervistato (ricordiamo che per la *Professione* del padre non era prevista la codifica a posteriori) ha permesso di effettuare un'analisi giornaliera per intervistatore delle *performance* della codifica assistita e ha fornito elementi utili alla valutazione dell'operato dei rilevatori, consentendo di compiere, nel corso della rilevazione, interventi *ad hoc* su coloro che mostravano particolari difficoltà nella gestione delle operazioni di codifica.

### 10.2.1 La durata della codifica della Professione

Iniziamo con l'analizzare le prestazioni del modulo di codifica assistita in termini di tempo medio necessario alla codifica (cfr. tabella 10.4). La codifica della *Professione* dei laureati ha richiesto in media 47 secondi, contro i 40 secondi necessari per codificare le *Professioni* dei diplomati. Le *Professioni* dei padri, invece, hanno richiesto tempi più brevi e praticamente identici nelle due indagini: 37 secondi per i padri dei laureati, 38 secondi per quelli dei diplomati.

La più elevata durata della codifica delle *Professioni* dei laureati può essere imputata ad una maggiore presenza tra i laureati di *Professioni* specialistiche e in generale più complesse da codificare, a volte definite anche con termini stranieri, e non sempre direttamente riconducibili a *Professioni* presenti nella Classificazione.



Tabella 10.4 – Indagini sulla Transizione istruzione-lavoro – La durata della codifica della Professione (secondi)

	<i>Professione Intervistato</i>	<i>Professione Padre</i>
<i>Indagine Laureati</i>	47''	37''
<i>Indagine Diplomati</i>	40''	38''

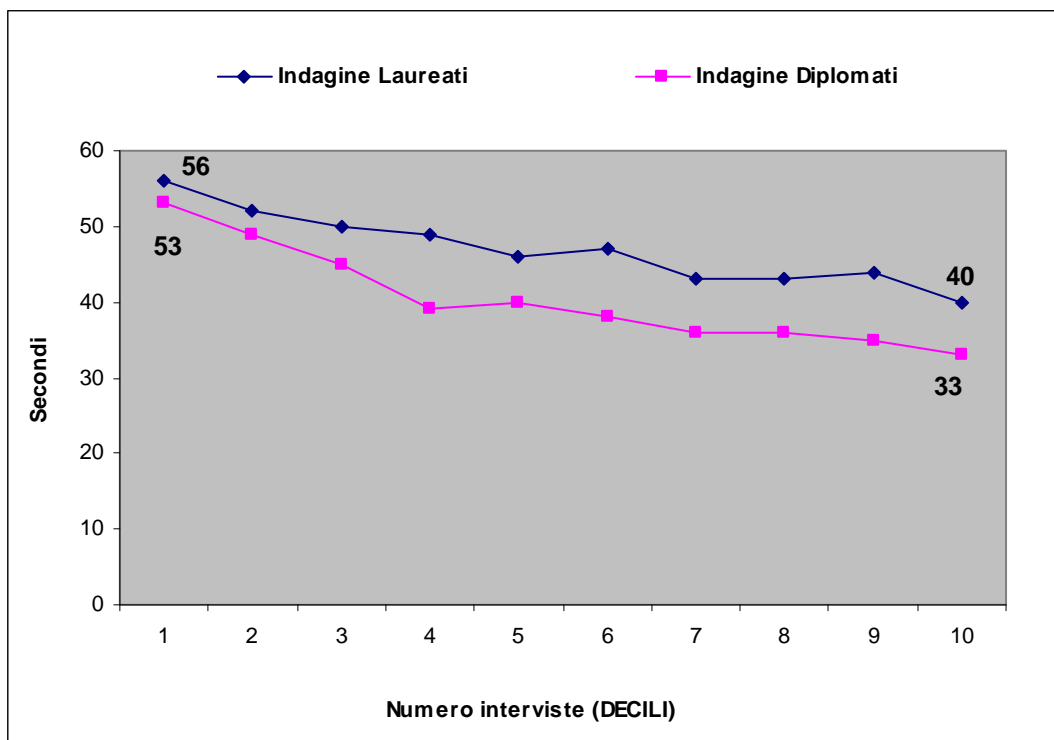
La distribuzione delle interviste secondo la durata della codifica mostra come quasi la metà delle *Professioni* dei diplomati (44,4%) siano state codificate in un tempo piuttosto breve (entro i 25 secondi); tale percentuale scende al 37,5% per le *Professioni* dei laureati (cfr. tabella 10.5). Per contro, le *Professioni* che hanno richiesto tempi più lunghi (oltre i 75 secondi) sono il 12% circa tra quelle dei diplomati e il 16,6% tra quelle dei laureati. Per quanto riguarda, invece, le *Professioni* dei padri, in entrambe le Indagini quasi la metà sono state codificate entro i 25 secondi, mentre solo una su dieci ha richiesto più di un minuto e 15 secondi.

In entrambe le rilevazioni i tempi di codifica della *Professione* degli intervistati sono stati più elevati nella fasi iniziali (primi decili di interviste), durante le quali i rilevatori dovevano ancora acquisire l'esperienza necessaria ad un efficiente utilizzo del modulo di codifica assistita (Figura 10.1). Nella prima parte di interviste, infatti, i tempi medi di codifica sono stati di poco inferiori al minuto (56 secondi per i laureati e 53 per i diplomati) e tuttavia sempre più bassi di quelli riscontrati nel corso dei test preliminari sulla codifica on line che, lo ricordiamo, avevano superato il minuto. Con il procedere delle due Indagini il tempo impiegato per la codifica è progressivamente diminuito scendendo, nell'ultima tranche di interviste, fino a 40 secondi in media nell'Indagine sui laureati e a 33 secondi nell'Indagine sui diplomati.

Tabella 10.5 – Indagini sulla Transizione istruzione-lavoro – Distribuzione percentuale delle interviste per durata della codifica della Professione

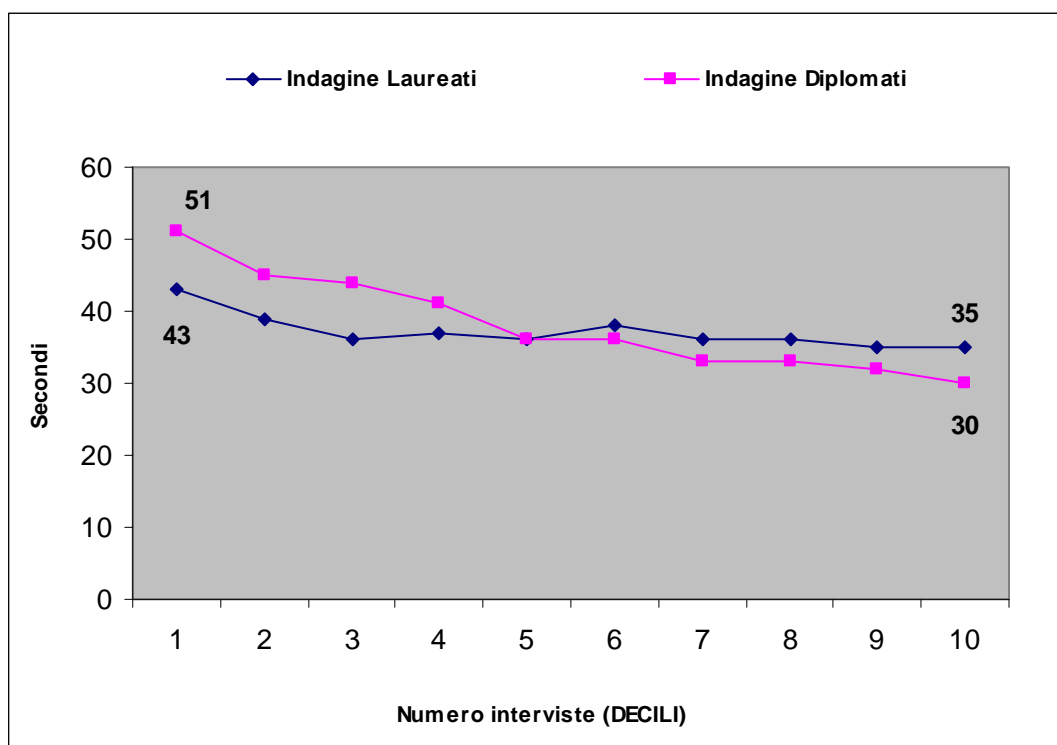
<b>SECONDI</b>	<i>Laureato</i>	<i>Padre del Laureato</i>	<i>Diplomato</i>	<i>Padre del Diplomato</i>
<i>Fino a 25''</i>	37,5%	49,2%	44,4%	49,5%
<i>Da 26'' a 45''</i>	28,2%	26,9%	26,7%	23,8%
<i>Da 46'' a 75''</i>	17,7%	14,1%	17,1%	15,3%
<i>Oltre 75''</i>	16,6%	9,8%	11,8%	11,4%
<b>Totale</b>	100%	100%	100%	100%

Figura 10.1 – La durata della codifica della Professione dell'intervistato nel corso dell'Indagine (secondi)



La codifica della *Professione* del padre ha richiesto, come si è visto, tempi analoghi nelle due Indagini. Nell'Indagine sui diplomati si è partiti da tempi più elevati nella prima parte (51 secondi di media) per poi scendere fino a una media di 30 secondi nelle ultime interviste (Figura 10.2). In quella sui laureati, invece, già nelle fasi iniziali si sono registrati tempi piuttosto bassi (43 secondi) che sono poi ulteriormente scesi, anche se in modo meno marcato, fino ad attestarsi su valori pressoché costanti nella seconda metà della rilevazione.

Figura 10.2– La durata della codifica della Professione del padre dell'intervistato nel corso dell'indagine (secondi)



di)

### 10.2.2 La codifica della Professione in corso d'intervista e a posteriori

Nel corso delle 25.510 interviste complete realizzate nel 2004 per l'Indagine sull'inserimento professionale dei laureati, la variabile *Professione* è stata rilevata e codificata in 42.780 casi: 17.813 per le attività professionali svolte dai laureati e 24.967 per le occupazioni dei padri.

Nell'Indagine sui percorsi di studio e di lavoro dei diplomati, in considerazione del minor numero di interviste e della maggior presenza di intervistati che non lavoravano al momento della rilevazione, la variabile *Professione* è stata rilevata in un numero minore di casi, pari a 30.720: 11.039 relativi ai diplomati e 19.681 concernenti i padri (cfr. tabella 10.6).

Tabella 10.6: Numero di Professioni dell'intervistato e del padre rilevate e codificate

PROFESSIONE	Indagine sui laureati Occorrenze	Indagine sui diplomati Occorrenze
<i>Intervistato</i>	17.813	11.039
<i>Padre</i>	24.967	19.681
<b>Totale</b>	<b>42.780</b>	<b>30.720</b>

Nell'Indagine sui diplomati sono state rilevate sia la *Professione* dell'intervistato che quella del padre nel 52,6% dei casi, mentre nell'Indagine sui laureati la doppia rilevazione ha riguardato il 68,3% delle interviste, essendo più alto tra i laureati il numero degli occupati a tre anni dal conseguimento del titolo rispetto ai diplomati, la gran parte dei quali risulta impegnata negli studi universitari (cfr. tabella 10.7). Nell'Indagine sui diplomati sono stati molti i casi di rilevazione e codifica di una sola *Professione* (45,3% contro il 31% dell'Indagine sui laureati), attività svolta quasi sempre dal padre; infine nello 0,6% delle interviste ai laureati e nel 2,1% di quelle realizzate ai diplomati non è stata raccolta alcuna *Professione* (alcuni casi riguardano diplomati o laureati che nel 2001 hanno conseguito il secondo diploma o la seconda laurea, collettivi di minor interesse rispetto ai neodiplomati e ai neolaureati, cui vengono richieste le sole informazioni anagrafiche).

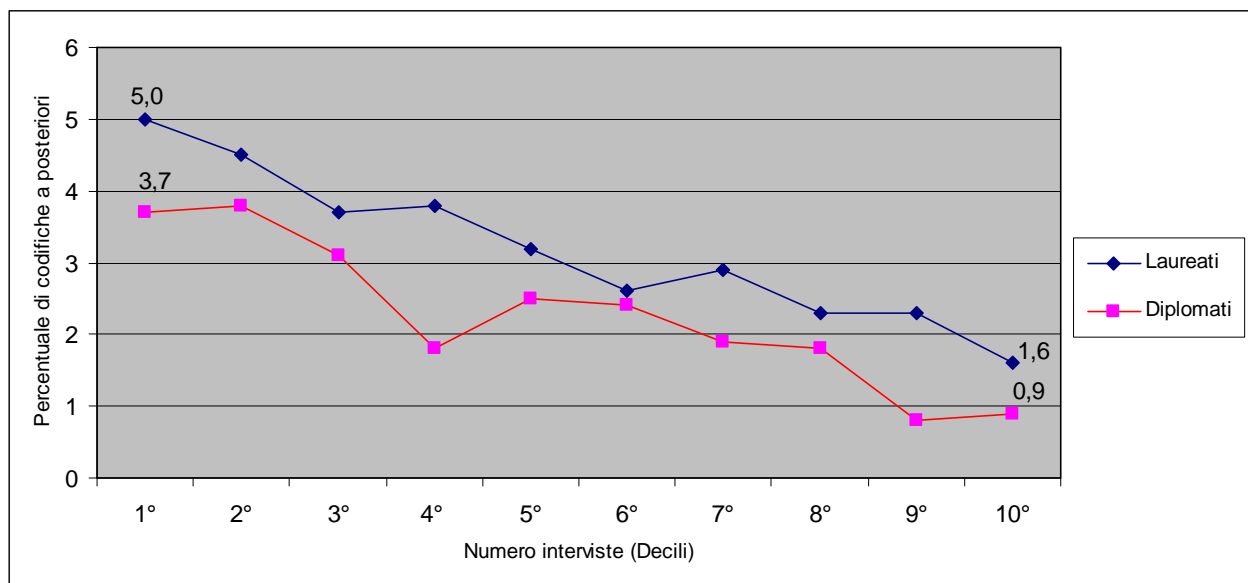
Tabella 10.7: Numero di Professioni rilevate e codificate per intervista

<i>PROFESSIONI RILEVATE PER INTERVISTA</i>	<i>Indagine sui laureati Occorrenze</i>	<i>%</i>	<i>Indagine sui diplomati Occorrenze</i>	<i>%</i>
<i>2 (intervistato e padre)</i>	17.434	68,3	10.743	52,6
<i>1 (solo padre)</i>	7.533	29,5	8.938	43,8
<i>1 (solo intervistato)</i>	379	1,5	296	1,5
<i>nessuna</i>	164	0,6	431	2,1
<b><i>Totale interviste complete</i></b>	<b>25.510</b>	<b>100,0</b>	<b>20.408</b>	<b>100,0</b>

In entrambe le indagini, la maggior parte delle *Professioni* svolte dall'intervistato sono state codificate nel corso dell'intervista, in particolare nel 96,8% dei casi nell'Indagine relativa ai laureati e nel 97,8% dei casi nell'Indagine sui diplomati, mentre le *Professioni* svolte dal padre sono state tutte codificate durante l'intervista. Le *Professioni* dell'intervistato sono quindi state codificate a posteriori nel 3,2% delle interviste a laureati che risultavano occupati al momento dell'intervista e nel 2,2% dei casi per l'Indagine sui diplomati.

Il maggior ricorso alla codifica a posteriori registrato nella prima indagine, così come i tempi più lunghi per l'attività di codifica durante l'intervista, conferma una maggiore complessità delle *Professioni* svolte da molti laureati, che rendono talvolta più difficili le operazioni di codifica da parte degli intervistatori. In entrambe le rilevazioni il ricorso alla codifica al termine dell'intervista è stato più rilevante nella fase iniziale della raccolta dati (1°decile di interviste), durante la quale gli intervistatori erano ancora inesperti nell'utilizzo del modulo di codifica assistita delle *Professioni*. Si è partiti da tassi di codifica a posteriori pari al 5% per l'Indagine sui laureati e al 3,7% per i diplomati, per scendere rispettivamente all'1,6% e allo 0,9% nell'ultima tranche di interviste, momento in cui gli intervistatori hanno affinato la tecnica di rilevazione e codifica e sono riusciti a risolvere un maggior numero di casi nel corso dell'intervista (Figura 10.3).

Figura 10.3: Incidenza percentuale della codifica a posteriori per decile di interviste



I lievi rialzi osservati in corrispondenza dei decili centrali sono dovuti all'ingresso di nuovi operatori dotati di minor esperienza nelle operazioni di rilevazione e codifica della *Professione* e dunque più portati a ricorrere alla codifica della *Professione* a posteriori.

### 10.2.3 Il dettaglio della codifica delle Professioni rilevate

L'analisi del livello di dettaglio della codifica della *Professione* dell'intervistato, ossia il numero di digit presenti nel codice assegnato, evidenzia, per entrambe le rilevazioni, un incremento della percentuale di *Professioni* codificate col massimo dettaglio rispetto alla precedente edizione: si è infatti partiti da una percentuale di *Professioni* codificate a 4 digit che si attestava attorno al 96-97% nelle rilevazioni condotte nel 2001, per arrivare alla quasi totalità dei casi (99,8%) nelle Indagini svolte nel 2004 (cfr. tabella 10.8).

Tabella 10.8: Livello di dettaglio della codifica della Professione dell'intervistato  
Codifiche a 4 digit

PROFESSIONE	Indagine 2004	Indagine 2001
Laureato	99,8%	96,8%
Diplomato	99,8%	96,2%

Malgrado vi fosse una minor esigenza di raccogliere la *Professione* svolta dal padre con un elevato livello di dettaglio, anche tale *Professione* è stata rilevata e codificata a 4 digit in più del 90% delle interviste e precisamente: nel 90,9% dei casi relativi ai padri dei laureati e nel 94% per i padri dei diplomati, con un incremento di 2-3 punti percentuali rispetto alle rilevazioni realizzate nel 2001 (cfr. tabella 10.9).

Tabella 10.9: Livello di dettaglio della codifica della Professione del padre  
Codifiche a 4 digit

PROFESSIONE	Indagine 2004	Indagine 2001
Padre del laureato	90,9%	88,8%
Padre del diplomato	94,0%	90,8%

### 10.3 La qualità della codifica delle Professioni rilevate

Al termine della fase di rilevazione si è voluta effettuare un'analisi di qualità della *Professione* svolta dall'intervistato. Ciò è stato realizzato esaminando un campione pari al 35-40% delle interviste realizzate a soggetti occupati al momento della rilevazione, per studiare la coerenza tra la descrizione della *Professione* digitata e il codice selezionato. Ciascuna *Professione* è stata quindi nuovamente codificata da personale esperto interno, sia con il ricorso al modulo di codifica assistita sviluppato in BLAISE sia tramite la consultazione del volume relativo alla Classificazione delle *Professioni*.

Il confronto tra la nuova codifica così ottenuta e quella precedentemente attribuita dagli operatori CATI nel corso dell'intervista ha dato luogo a tre possibili risultati: 1. perfetta corrispondenza tra descrizione e codifica associata dall'intervistatore – ossia alla *Professione* raccolta è associabile la sola codifica correttamente attribuita; 2. compatibilità tra descrizione e codifica associata nel corso dell'intervista – alla *Professione* rilevata in modo incompleto è stata attribuita una delle possibili codifiche associabili; 3. nessuna corrispondenza tra descrizione e codifica – l'intervistatore ha codificato la *Professione* in modo errato oppure non ha digitato in modo corretto la descrizione.

L'analisi ha avuto il seguente esito: per l'Indagine sui laureati, sulla base di 6.261 casi osservati nel 98,6% dei casi la *Professione* rilevata è risultata codificata in modo corretto, nell'1,1% delle interviste alla *Professione* rilevata, è stato attribuito un codice comunque compatibile con la descrizione riportata (ad esempio alla professione 'professore' è stata associata una codifica relativa ad un insegnante di scuola media o superiore) e nello 0,2% delle interviste non vi è alcuna corrispondenza tra la descrizione e la codifica (ad es. alla descrizione 'barista' è stato associato un codice appartenente al Grande Gruppo 1) (cfr. tabella 10.10).

Tabella 10.10: Coerenza tra descrizione della Professione dell'intervistato e codifica della Professione  
Indagine sui laureati Indagine sui diplomati

<i>CODIFICHE</i>	<i>Occorrenze</i>	<i>%</i>	<i>Occorrenze</i>	<i>%</i>
<i>Corrette</i>	6.172	98,6	3.932	98,7
<i>Compatibili</i>	71	1,1	36	0,9
<i>Non corrette</i>	18	0,3	16	0,4
<b><i>Totale</i></b>	<b>6.261</b>	<b>100,0</b>	<b>3.984</b>	<b>100,0</b>

Risultati analoghi si osservano nell'Indagine sui diplomati: sulla base di 3.984 interviste riguardanti diplomati che al momento dell'intervista risultavano occupati, nel 98,7% dei casi la *Professione* rilevata è stata codificata in modo corretto, nello 0,9% alla *Professione* rilevata è stato attribuito un codice compatibile con la descrizione riportata e nello 0,4% dei casi non vi è alcuna corrispondenza tra la descrizione e il codice.

All'interno dei questionari utilizzati per le due rilevazioni, oltre ai quesiti relativi alla *Professione* svolta dall'intervistato e dal padre, sono presenti anche domande concernenti la posizione nella *Professione* di entrambi i soggetti e il titolo di studio posseduto dal padre. Per uno studio più approfondito sulla qualità della codifica della *Professione* (svolta dall'intervistato e dal padre) sono in corso analisi sulla coerenza tra la descrizione della *Professione* fornita dagli intervistati, la relativa codifica attribuita, la posizione nella *Professione* e, per il padre, il titolo di studio posseduto.

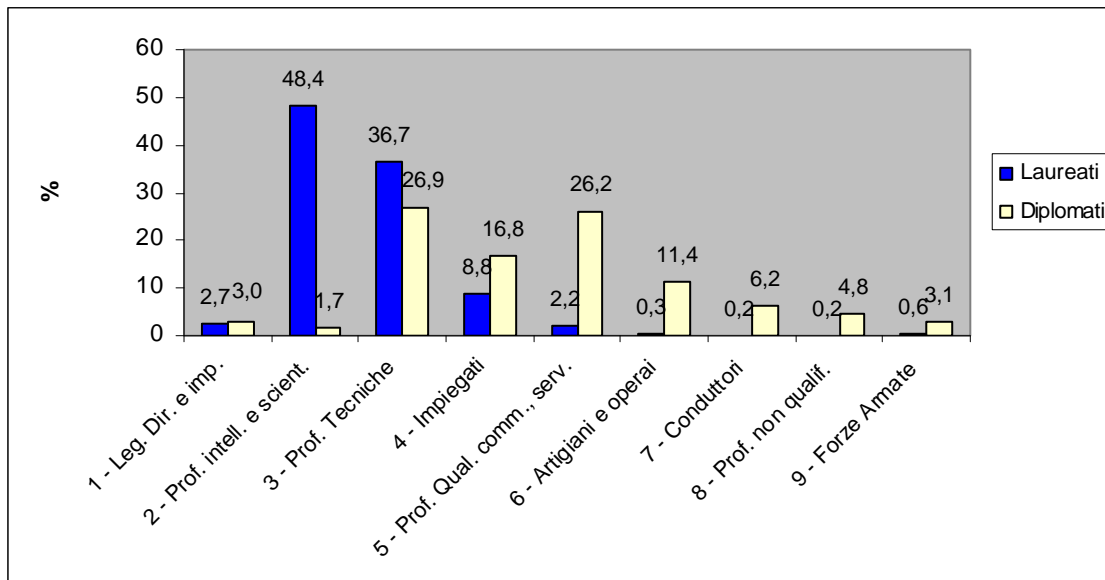
#### 10.4 Analisi delle Professioni rilevate per Grande Gruppo

La distribuzione delle *Professioni* svolte dagli intervistati per Grande Gruppo evidenzia che:

- tra i laureati prevalgono quelle appartenenti al Grande Gruppo 2 (professioni intellettuali e scientifiche) che rappresentano il naturale sbocco professionale dei laureati, per lo svolgimento delle quali è richiesto quasi sempre il possesso di un titolo di studio universitario, e al Grande Gruppo 3 (professioni tecniche) nel quale si collocano i laureati che non riescono, almeno nella prima fase del processo di transizione, a valorizzare appieno il titolo universitario conseguito (*over education*);
- tra i diplomati prevalgono invece quelle concernenti il Grande Gruppo 3, ossia professioni tecniche per lo svolgimento delle quali è generalmente richiesto il diploma di scuola secondaria superiore, e i Grandi Gruppi 4 (impiegati esecutivi) e 5 (professioni qualificate nei servizi). Una rilevante quota di diplomati (22,4% dei casi) svolge *Professioni* appartenenti ai Grandi Gruppi 6, 7 e 8. La loro presenza all'interno del Grande Gruppo 1 è giustificata dal fatto che per alcune delle *Professioni* appartenenti a questo Gruppo non è necessario possedere un titolo universitario (esempio: imprenditore o dirigente di azienda privata); stesso discorso, anche se in maniera più limitata, vale per le *Professioni* svolte all'interno del Grande Gruppo 2, in particolare nella categoria 2.5.5 (specialisti in discipline artistico-espressive: pittori, attori, ballerini, cantanti) cui sono da sommare i casi relativi a diplomati che a distanza

di tre anni dal conseguimento del titolo di scuola secondaria superiore hanno terminato con successo gli studi universitari e svolgono una *Professione* d'elevato profilo (Figura 10.4).

Figura 10.4: Professione svolta dall'intervistato per Grande Gruppo



La distribuzione delle *Professioni* svolte dai padri degli intervistati per Grande Gruppo evidenzia, per i padri dei laureati, una più elevata presenza nei primi tre Grandi Gruppi, che ammonta al 53,2% (12,6% nel G.G.1, 18,4% nel G.G. 2, 22,2% nel G.G. 3); tra i padri dei diplomati tale valore è pari al 36,3% e risulta inferiore rispetto a quello relativo ai Grandi Gruppi 6, 7 e 8, concernenti le *Professioni* svolte dagli artigiani, dagli operai qualificati e dai lavoratori senza specifica qualificazione, ove raggiunge il 43,4% (Figura 10.5).

Come accennato in precedenza, i casi di codifica della *Professione* del diplomato effettuata al termine dell'intervista sono stati veramente contenuti: 245 pari al 2,2% delle *Professioni* rilevate.

Le maggiori difficoltà incontrate dagli intervistatori nelle operazioni di codifica hanno riguardato in particolar modo le *Professioni* appartenenti ai Grandi Gruppi 7 (conduttori di impianti e macchinari) 4,7%, 1 (legislatori, dirigenti e imprenditori) 3,9%, 8 (professioni non qualificate) 3,4% e 6 (artigiani e operai specializzati) 3,1% (Figura 10.6).

Si tratta comunque di percentuali contenute, legate alla presenza all'interno di questi Grandi Gruppi di *Professioni* più complesse da codificare che necessitano di una conoscenza più approfondita della Classificazione delle professioni da parte degli intervistatori.

Anche per l'Indagine sui laureati la codifica a posteriori della *Professione* ha riguardato poche centinaia di interviste: 594, pari al 3,2% delle *Professioni* rilevate.



Figura 10.5: Professione svolta dal padre dell'intervistato per Grande Gruppo

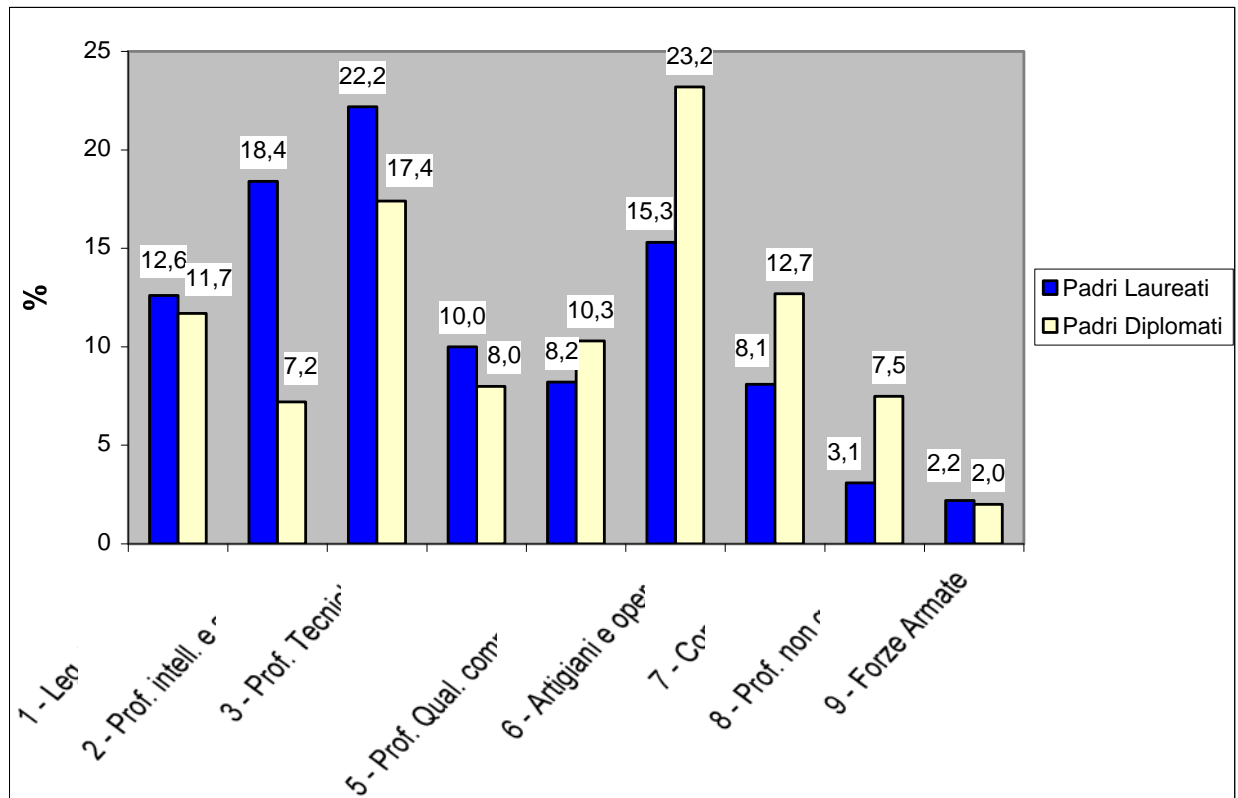
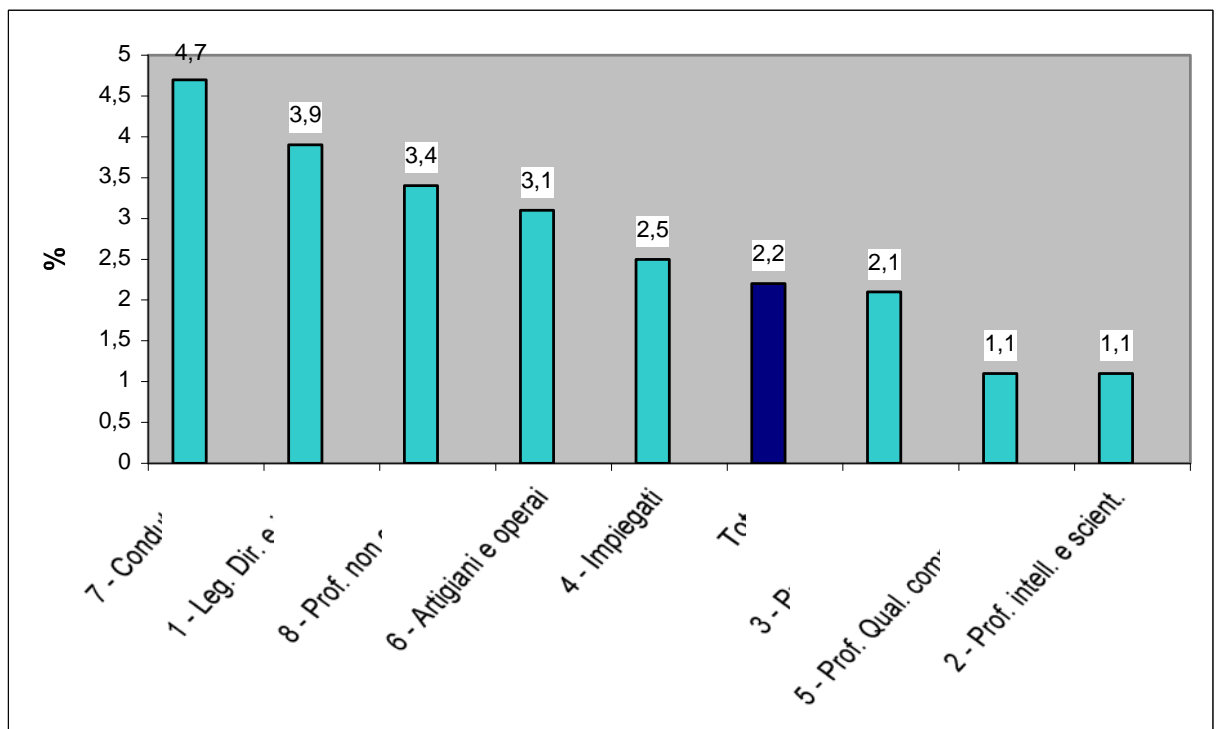
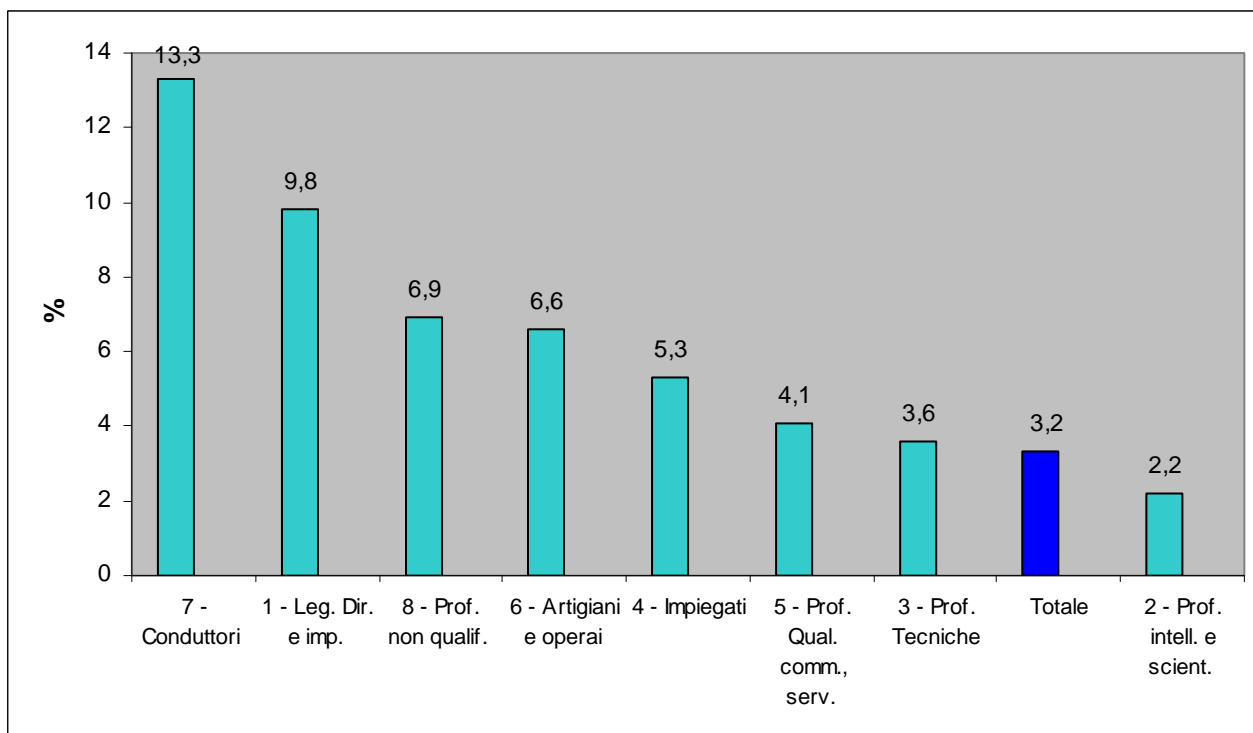


Figura 10.6: Incidenza della codifica a posteriori della Professione del diplomato per Grande Gruppo



Come riscontrato nell'Indagine sui diplomati, si osservano percentuali più elevate di codifica a posteriori nei Grandi Gruppi 1, 7, 8 e 6. Nell'Indagine sui laureati tali percentuali risultano anche più elevate in quanto la scarsa presenza di laureati occupati in *Professioni* appartenenti a tali Gruppi ha comportato una minor possibilità per gli intervistatori di affinare le capacità di codificare tali tipi di *Professioni* (Figura 10.7).

Figura 10.7: Incidenza della codifica a posteriori della Professione del laureato per Grande Gruppo



Le maggiori difficoltà relative alla codifica delle *Professioni* appartenenti ad alcuni Grandi Gruppi, evidenziate dal più frequente ricorso alla codifica a posteriori, sono confermate anche dai relativi tempi di codifica più alti nel corso dell'intervista. Le *Professioni* svolte dagli intervistati appartenenti al Grande Gruppo 7 sono quelle più complesse da codificare infatti richiedono poco più di un minuto d'intervista, (61 secondi per quelle svolte dai diplomati e 74 secondi per le attività svolte dai laureati), seguono quelle concernenti i G.G. 1, 6 (per i laureati) e 4; decisamente più agevole risulta essere la codifica delle *Professioni* svolte nell'ambito delle Forze Armate (G.G. 9) e delle attività qualificate nei servizi svolte dai diplomati (G.G. 5) che impegnano l'intervistatore per circa mezzo minuto d'intervista (cfr. tabelle 10.11 e 10.12).

Analoghe tendenze si osservano per la rilevazione della *Professione* del padre ove le *Professioni* dei G.G. 4, 1 e 7 rappresentano gli ostacoli più difficili da superare e richiedono in media tra i 45 e i 50 secondi d'intervista.

Tabella 10.11: Tempi di codifica della Professione nel corso dell'intervista per Grande Gruppo  
Indagine sui percorsi di studio e di lavoro dei diplomati

Grande Gruppo	Diplomato	Padre
1 - Legislatori, dirigenti e imprenditori	50"	49"
2 - Professioni intellettuali, scientifiche e di elevata specializzazione	38"	29"
3 - Professioni tecniche	39"	40"
4 - Impiegati	49"	50"
5 - Professioni qualificate nelle attività commerciali e nei servizi	29"	29"
6 - Artigiani, operai specializzati e agricoltori	40"	32"
7 - Conduttori di impianti e operai semiqualeficati addetti a macchinari fissi e mobili	1'01"	46"
8 - Professioni non qualificate	48"	35"
9 - Forze Armate	31"	29"
<b>Totale</b>	<b>40"</b>	<b>38"</b>

Tabella 10.12: Tempi di codifica della Professione nel corso dell'intervista per Grande Gruppo  
Indagine sull'inserimento professionale dei laureati

Grande Gruppo	Laureato	Padre
1 - Legislatori, dirigenti e imprenditori	1'12''	46''
2 - Professioni intellettuali, scientifiche e di elevata specializzazione	40''	26''
3 - Professioni tecniche	55''	39''
4 - Impiegati	1'07''	50''
5 - Professioni qualificate nelle attività commerciali e nei servizi	53''	28''
6 - Artigiani, operai specializzati e agricoltori	1'09''	34''
7 - Conduttori di impianti e operai semiqualeficati addetti a macchinari fissi e mobili	1'14''	46''
8 - Professioni non qualificate	47''	37''
9 - Forze Armate	28''	26''
<b>Totale</b>	<b>47''</b>	<b>37''</b>

## Conclusioni

Le innovazioni introdotte per la rilevazione, codifica e monitoraggio della *Professione* hanno permesso di ottenere ottimi risultati in termini di qualità dei dati e di tempi di rilevazione. L'esperienza effettuata potrebbe essere estesa a tutte le Indagini CATI su popolazione sviluppate con la '*strategia in-house*': fermi restando i vincoli che sottendono l'utilizzo della codifica assistita di una variabile, ossia complessità della Classificazione e lunghezza dell'intervista, lo strumento *software* potrebbe essere implementato all'interno di un qualsiasi questionario elettronico sviluppato in BLAISE e associato ad una fase di formazione e di monitoraggio di entità e peso variabili in funzione dell'importanza rivestita dalla variabile *Professione* nell'ambito delle Indagini. Il vantaggio nell'utilizzo della codifica assistita delle *Professione* risulterebbe anche in un'integrazione continua della base informativa: in funzione delle varie tipologie di rispondenti, il dizionario potrebbe arricchirsi di nuove *Professioni* emergenti tra le quali anche quelle afferenti a particolari segmenti di popolazione.

## Bibliografia

- ACTRV3 User Guide (1998), Statistics Canada
- Blaise Developer's Guide (1999), Statistics Netherlands
- C. Castagnaro, M. Lo Conte, S. Macchia, M. Murgia (2003), 'Una soluzione in house per le indagini CATI: il caso della Indagine Campionaria sulle Nascite' *Documenti Istat*, n.8 2003
- Cutillo A., Pace S., Parisi V., Rottino F.M. (2004) "Reorganising the process of producing data to improve the quality of the sample surveys: the survey on the study and work decisions of the qualified in the upper secondary school" *European conference on quality and methodology in Official Statistics, Mainz*, 24-26 maggio 2004
- De Angelis R., Macchia S. and Mazza L. (2000), Applicazioni sperimentali della codifica automatica: analisi di qualità e confronto con la codifica manuale, *Istat Quaderni di ricerca – Rivista di statistica Ufficiale*, 1, 29-54
- Istat Classificazione delle Professioni, *Metodi e norme*– nuova serie n. 12 edizione 2001
- Istat (2003): I laureati e il mercato del lavoro – Inserimento professionale dei laureati - Indagine 2001, *Collana Informazioni*.
- Istat (2004): I laureati e lo studio – Inserimento professionale dei laureati - Indagine 2001, *Collana Informazioni*.
- Jones N., Lewis D. (2004) Developing a handbook on improving quality by analysis of process variables *European Conference on quality and methodology in Official statistics - Q2004, Mainz, May 24-26*
- Lyberg L. et al. (eds) (1997), *Survey Measurement and Process Quality*, New York, Wiley.
- Macchia S. (2001), Integration of sources to build a dictionary for Automated Coding of Industry, *Riunione scientifica del gruppo SIS Classificazione e Analisi dei dati (Cladag)*, Palermo, 5-6 luglio 2001
- Macchia S., L. Mazza, P. Scalisi (2003) 'Which is your Occupation? It is not so easy to answer' *Riunione scientifica del gruppo SIS Classificazione e Analisi dei dati (CLADAG)*, Bologna 22-24 settembre 2003
- Macchia S., S. Mastroluca (2004) "The automatic coding process in the 2001 Italian General Population Census: efficacy and quality", *European conference on quality and methodology in Official Statistics, Mainz*, 24-26 maggio 2004
- Micali A., Ungaro P. (1998) Il sistema di indagini Istat sulla transizione scuola-lavoro. *SIEDS Rivista italiana di economia e statistica*, LII, 311-324.
- Murgia M., Pace S. Ungaro P., 2002, "Automated coding system: an application to the ISTAT school-to-work transition survey", in *Società Italiana di Statistica (2002), Atti della XLI Riunione scientifica*, Università di Milano Bicocca, 5-7 Giugno 2002
- Macchia S., Mazza L., Ungaro P. (2005), 'Dalla descrizione della professione al codice: esperienze di codifica assistita nelle rilevazioni telefoniche', *Giornate di studio della popolazione del gruppo SIS*, Padova 16-18 febbraio 2005
- Ryan T. P. (2000) *Statistical methods for quality improvement (Second edition)*, Wiley, New York.

Simeoni G., Grassia M. G. (in stampa) Strumenti per il monitoraggio del processo: l'uso delle carte di controllo, in *Il monitoraggio del processo e la stima dell'errore nelle indagini telefoniche. Applicazioni all'indagine sulla Sicurezza dei cittadini*, a cura di Muratore M. G., Signore M., Collana *Metodi e Norme*, ISTAT

Wenzowski M.J. (1988). ACTR – A Generalised Automated Coding System. *Survey Methodology*, vol. 14: 299-308

WESTAT, (1992) *Statistical thinking*. Rockville, Maryland. Manuale fornito dalla WESTAT per un corso ai TQM-pilots di Statistics Sweden. Documento interno