

Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea in Scienze Statistiche ed Economiche



La ricerca sociale nell'era digitale.
Il caso dell'International Social Survey Programme

Relatrice Prof.ssa Arjuna Tuzzi
Dipartimento di Filosofia, Sociologia, Pedagogia e Psicologia applicata

Laureando: Gianni Bregolin
Matricola N. 246688

Anno Accademico 2013/2014

INDICE

PREMESSA	1
1 – RIFLESSIONI E RIFRAZIONI METODOLOGICHE	11
1.1 L'IMPATTO DELLA TECNOLOGIA NEI DISEGNI D'INDAGINE.....	12
1.1.1 <i>Lo scenario tecnologico e i problemi inerenti il campionamento probabilistico</i>	12
1.1.1.1 Premessa.....	12
1.1.1.2 La convergenza degli strumenti di comunicazione.....	14
1.1.1.3 Il campionamento probabilistico: vittima del digital divide?.....	14
1.1.2 <i>Interazioni tra tecnologia, cultura di ricerca e idee sociali</i>	21
1.1.2.1 Universi liquidi e riponderazioni delle stime	22
1.1.2.2 Numerosità campionaria e omogeneità dei sub-universi tecnologici: il mito dei rispondenti fidelizzati	25
1.1.2.3 La credibilità della ricerca sociale: in bilico tra cultura aristotelica e metodo scientifico.....	29
1.1.2.4 Idee Sociali: la tecnologia e la privacy.....	31
1.1.3 <i>Ricerca sociale e governo del mutamento: una breve riflessione</i>	33
1.1.4 <i>Tecnologia e raccolta dati</i>	33
1.1.4.1 Unificazione dei supporti per la raccolta dei dati.....	34
1.1.4.2 Ibridazione delle tecniche di raccolta dati e ruolo del rilevatore.....	35
1.1.4.3 Lo strumento di rilevazione: da questionario a gestore di processo	40
1.2 LA PRODUZIONE DI INFORMAZIONE E LE TRASFORMAZIONI DEL COMPARTO DELLA RICERCA	47
1.2.1 <i>Premessa</i>	47
1.2.2 <i>La natura dei costi e l'evoluzione dei processi produttivi</i>	47
1.2.3 <i>Qualità nei processi di raccolta dei dati</i>	49
1.2.4 <i>Trasformazioni del comparto produttivo</i>	54
1.2.5 <i>Strutture centralizzate e tempi di rilevazione</i>	58
1.3 IL VALORE ECONOMICO DELLA COPERTURA CAMPIONARIA: IL "COSTO METODOLOGICO"	59
1.3.1 <i>Il confronto tra tecniche d'indagine: una proposta</i>	61
1.3.2 <i>Il virus degli spare-panel</i>	63
1.4 SFIDE ALLE SCIENZE SOCIALI	67
2 – L'INDAGINE SCIENTIFICA ISSP	71
2.1 ISSP: BREVE PRESENTAZIONE	72
2.2 VINCOLI DEL COMITATO SCIENTIFICO ISSP SULL'IMPIANTO DELL'INDAGINE.....	77
2.3 RICOGNIZIONE OPERATIVA PER LA PREDISPOSIZIONE DELL'INDAGINE IN ITALIA.....	80

2.3.1	<i>L'indagine pilota</i>	80
2.3.2	<i>La disponibilità di campioni probabilistici in Italia</i>	81
2.3.3	<i>Schema sintetico dei principali ostacoli e obiettivi</i>	81
2.4	IL PIANO DI RICERCA ISSP IN ITALIA: CENNI DELLA SOLUZIONE ADOTTATA	83
2.4.1	<i>Il piano di campionamento</i>	83
2.4.2	<i>Il piano operativo di rilevazione</i>	84
2.4.3	<i>Una possibile definizione dell'indagine ISSP in Italia</i>	86
3	– GEOGRAFIE CAMPIONARIE	89
3.1	ESPERIENZE INTERNAZIONALI: DAL CATI ALLE INDAGINI POSTALI	90
3.1.1	<i>Il CATI: storia di un eldorado perduto</i>	90
3.1.2	<i>Verso l'Address Based Sampling (ABS)</i>	96
3.1.2.1	Panel e campioni sperimentali istantanei: diversi strumenti per diversi obiettivi	97
3.1.3	<i>Pregi e difetti dei disegni d'indagine basati sugli indirizzi</i>	99
3.1.4	<i>ABS e Geocodifica</i>	101
3.1.4.1	La geocodifica per la selezione della lista campionaria	101
3.1.4.2	Qualità dei DSF e qualità del processo di geocodifica	103
3.2	REPLICABILITÀ DELLE INDAGINI ABS IN ITALIA	105
3.2.1	<i>Assenza dei file DSF e le fonti anagrafiche in Italia</i>	105
3.2.2	<i>Il Reverse Geocoding e la proposta di campionamento per l'Italia</i>	106
3.2.2.1	L'individuazione del civico dove recapitare il questionario	108
3.2.2.2	Lo sviluppo verticale delle aree urbane	110
3.2.3	<i>Prime riflessioni sulla strategia di campionamento sviluppata</i>	114
3.3	DETTAGLI E LINEE DI SVILUPPO DEL PIANO DI CAMPIONAMENTO ISSP ITALIA	116
3.3.1	<i>Primo stadio: la stratificazione per geotipologie</i>	116
3.3.2	<i>Secondo stadio: i punti di campionamento (comuni)</i>	117
3.3.3	<i>Limitazione del danno, dei costi e le diversità geotipologiche: un approfondimento sulla rete dei rilevatori</i>	119
3.3.4	<i>Terzo stadio: le famiglie contattate all'interno dei comuni sondati</i>	120
3.3.5	<i>Quarto stadio: i rispondenti</i>	120
3.4	IL CAMPIONAMENTO AREALE PER LE SCIENZE SOCIALI: LA PROPOSTA AMERICANA	121
3.4.1	<i>L'impatto della tecnologia sulle prassi e sulle teorie del campionamento</i>	121
3.4.1.1	La ricostruzione dell'universo di campionamento	123
3.4.1.2	Sull'indice di bontà dell'esperimento e sul significato ultimo dell'esperimento proposto	126
3.4.1.3	Sulla problematica dello sviluppo verticale delle realtà urbane	129

3.4.2	<i>Lo sviluppo teorico dei campionamenti areali.....</i>	130
3.4.2.1	<i>La selezione delle unità campionarie.....</i>	131
3.4.2.2	<i>La propagazione delle informazioni censuarie a livello micro-areale.....</i>	133
3.4.3	<i>Sintesi e spunti di riflessione sullo sviluppo della tecnologia geo-informativa.....</i>	135
4	– IL FIELDWORK E ALCUNI RISULTATI OTTENUTI.....	139
4.1	PREMESSA.....	140
4.2	I RILEVATORI.....	142
4.2.1	<i>La scheda di consegna e il ruolo dei rilevatori.....</i>	142
4.2.2	<i>Caratteristiche desiderate per il rilevatore.....</i>	145
4.2.3	<i>Metodo e costo del reclutamento dei candidati rilevatori.....</i>	146
4.2.4	<i>L'accreditamento e la selezione delle candidature.....</i>	147
4.2.5	<i>Alcuni aspetti legati alla formazione dei rilevatori.....</i>	150
4.3	IL COMPLETAMENTO DEL PROCESSO DI CAMPIONAMENTO.....	151
4.3.1	<i>Eleggibilità ed ineleggibilità dell'indirizzo.....</i>	151
4.3.2	<i>Individuazione dell'edificio.....</i>	152
4.3.3	<i>Individuazione del nucleo familiare: l'uso delle tavole casuali.....</i>	154
4.4	IL FIELDWORK E I RISULTATI OTTENUTI.....	155
4.4.1	<i>La consegna del questionario.....</i>	155
4.4.2	<i>Il controllo sistematico sull'attività del rilevatore.....</i>	157
4.4.3	<i>Il feedback verso il rilevatore.....</i>	159
4.4.4	<i>I solleciti alla compilazione.....</i>	162
4.5	L'ADESIONE ALL'INDAGINE.....	165
4.5.1	<i>Le caratteristiche delle residenze.....</i>	165
4.5.2	<i>Il ruolo del CAWI.....</i>	166
4.5.3	<i>Il digital divide e le geotipologie.....</i>	168
4.6	RELAZIONI TRA ADESIONE, MODALITÀ DI CONSEGNA E SOLLECITI TELEFONICI.....	171
4.6.1	<i>Azioni per aumentare l'adesione al progetto.....</i>	174
4.6.2	<i>Un patrimonio da valorizzare.....</i>	176
	CONCLUSIONI.....	179
	RIFERIMENTI BIBLIOGRAFICI.....	183
	RINGRAZIAMENTI.....	189

Premessa

In un periodo in cui sembra inarrestabile l'incremento dei mezzi e dei modi con cui le persone possono entrare in contatto tra loro, la ricerca empirica in ambito sociale sembra più che mai in crisi: privata, innanzitutto, di un'efficace strumento con il quale svolgere le proprie indagini scientifiche. Per oltre un trentennio la tecnica CATI (*Computer Assisted Telephone Interviewing*) ha rappresentato il cavallo di battaglia sul quale tutti, società private e istituti di ricerca pubblici, sono saliti per condurre le proprie indagini presso la popolazione. Veloce, robusto e controllabile, il CATI è entrato precipitosamente in crisi nell'ultimo decennio, ovvero da quando l'accelerazione tecnologica nella comunicazione ha infranto la barriera che separava la grande massa delle persone da quelle nicchie di soggetti che per studio o lavoro s'erano addentrati molti anni fa nel mondo della tecnologia digitale.

A trent'anni circa dalla reale disponibilità dei primi *personal computer* (pc), si è oggi in presenza di reti di connessione *wireless* che permettono l'utilizzo di telefoni (gli *smartphone*) a volte più potenti (in termini computazionali e di memoria) di molti pc ancora presenti nelle aziende o nelle nostre case. Una rivoluzione che ha moltiplicato, come in un'esplosione, i nodi di connessione del grafo con il quale è possibile rappresentare le strutture fisiche di comunicazione. Il contatto tra gli individui è diventato sempre meno indiretto e sempre più individuale. I mezzi di comunicazione non fanno quasi più riferimento a strutture organizzate (la famiglia, l'ufficio, ecc.), ma sono sempre più personali e, per ciascuno, spesso ridondati. È così che il telefono fisso (dalle cabine sparse nella città al telefono di casa o dell'ufficio) è pressoché scomparso¹ o reso immateriale tramite la tecnologia VoIP (*Voice over Internet Protocol*). Non è nomade solo il telefono cellulare; lo è potenzialmente anche il numero di telefono fisso. Il concetto stesso di relazione tra numero/prefisso e

¹ In termini relativi rispetto al totale degli strumenti di comunicazione.

territorio è retaggio del passato e della tecnologia analogica ad esso collegata. Oggi, è *nomadic* il *worker* perché lo è l'insieme degli strumenti che usa. Un pc può essere acceso oggi a Londra e domani a Caracas, e si è sempre raggiungibili con lo stesso numero d'ufficio con cui ci si invita, di stanza in stanza, a prendere un caffè al bar di fronte². Una rivoluzione, si è detto, che ha digitalizzato anche le forme stesse della comunicazione. La classica telefonata è diventata una videochiamata; una lettera o una fattura è ora una mail con i relativi allegati; una battuta tra amici è possibile scambiarla scegliendo tra una vasta gamma di canali e opzioni: un sms, una chat, un *tweet*, ecc., così da farla sentire a chi ci "segue", scambiarla in privato o, più semplicemente, lanciarla nell'affollato vuoto della rete. In questo mondo ipercinetico, è quasi paradossale prendere atto che il principale problema della ricerca campionaria in ambito sociale sia quello della copertura, ossia dell'incapacità di entrare in contatto con tutto l'universo di riferimento sul quale si indaga.

Come in ogni rivoluzione, anche quella digitale apre le porte a possibili derive che, come spesso accade, sembrano rappresentare l'adeguata risposta ai tumultuosi eventi che accadono quotidianamente. Questo, infatti, è il tempo in cui alla richiesta di essere sempre più veloci ed economici nella raccolta delle informazioni, gli istituti di ricerca contrappongono *pseudo panel* o, meglio, *spare-panel*³ di individui, spacciati per campioni, da consultare per qualsiasi argomento. Nella quasi totalità dei casi, questi "*panel*" sono composti da soggetti che, spesso pagati e di certo non appagati solo dal loro inconsapevole ruolo di testimoni, si allineano prodighi a trovare sempre una risposta tra quelle che il

² Attraverso il *cloud computing*, inoltre, molta parte dell'insieme strumentale del proprio lavoro diventa sempre e ovunque disponibile.

³ Si è voluto coniare il termine *spare-panel* per mettere in evidenza il concetto di "scorta" associato a quello di "panel", che a sua volta identifica il tipico campione (un lista di rispondenti sostanzialmente stabile nel tempo) utilizzato nelle indagini longitudinali finalizzate a misurare l'evoluzione nel tempo delle caratteristiche di interesse. Un "campione di scorta", da intendersi con un significato quasi "dispregiativo", che nulla ha a che fare con la ripetizione dell'esperimento probabilistico di campionamento iniziale (*continuous sampling*), ma che al contrario sottende l'abuso sistematico di una "lista di sostituzione".

ricercatore ha inserito nel questionario di turno⁴. È questo uno dei tanti rischi in cui si potrebbe incappare usando in modo incontrollato le indagini CAWI (*Computer Assisted Web Interviewing*), largamente declamate come modello di modernità ed efficienza e quasi mai spiegate come schiera di assoldati individui a disposizione degli istituti di ricerca. Altre volte, come succede in molte aziende che si occupano di indagini CATI, chi risponde ad un questionario telefonico è inserito a sua insaputa nel contenitore dei numeri di telefono dei disponibili/profilati (Di Gioia, 2009), e ricontattato per un'altra indagine qualora ci sia la necessità di rincalzare alcune quote campionarie (spesso composte da giovani, maschi, ancor meglio se appartenenti alle fasce attive della popolazione)⁵. A questi “signori” della ricerca, bisognerebbe ricordare che nemmeno la disponibilità gratuita è una variabile statisticamente indifferente (indipendente) rispetto ai risultati ottenuti. È noto, ad esempio, che nelle indagini elettorali ciò rappresenta (per motivi del tutto da approfondire) un carattere di differenziazione tra la popolazione che si dichiara di centro-destra rispetto a quella di centro-sinistra, quest'ultima generalmente sempre troppo presente nel campione effettivamente sondato. Se lo è la disponibilità gratuita, figuriamoci quella remunerata! Ma guardando al panorama complessivo, questo non rappresenta neppure il male peggiore se, come accade, gli istituti di ricerca sono spesso a doversi difendere da accuse circa la loro disponibilità ad essere promotori di determinati punti di vista, piuttosto che capaci e attenti nel trovare il modo migliore per cogliere e misurare un fenomeno. In ogni caso, anche quando non è l'obiettivo di un

⁴ Il fatto che l'intervistato venga ricompensato economicamente per la sua partecipazione all'indagine può amplificare ancor più gli errori riconducibili all'accondiscendenza potenziale tra il soggetto e, in questo caso, i punti di vista con i quali si sono sviluppati gli strumenti d'indagine. Tale fenomeno, noto in letteratura (Roccatò, 2003), è stato riscontrato anche relativamente alle interviste telefoniche, nelle quali intervistatore e intervistato entrano in contatto per la prima volta e con una modalità che limita la piena espressività/comunicazione tra gli individui (il telefono). L'intonazione vocale o il modo con il quale vengono poste le domande, possono far nascere l'idea che esistano delle aspettative da parte dell'intervistatore sui temi trattati, attese che vengono assecondate dall'intervistato.

⁵ Un uso massiccio e ripetitivo degli *spare-panel* comporta una potenziale espansione incontrollata della distorsione campionaria – effetto *attrition* relativo ai panel veri e propri (Lee, 2003) – già di per sé presente a causa di una disponibilità all'intervista non omogenea tra i diversi gruppi sociali.

indebita profitto ad inanellare queste perle d'opera, si inseriscono guizzi di curiosità (più che di rigore metodologico) che dovrebbero essere accuratamente declinati come tali piuttosto che spacciati come valide alternative ad un metodo di ricerca scientifico. Nelle recenti elezioni politiche del 2013, ad esempio, tutti gli istituti di ricerca hanno molto sottovalutato quanti avrebbero votato per il Movimento 5 Stelle. Qualcuno ha azzardato l'ipotesi che tutto ciò dipenda dal fatto che si fa solo ricerca telefonica e poco spazio sia dato alla rete. Ma esiste un modo per interrogare scientificamente/correttamente la rete? Recenti analisi svolte da un gruppo di ricerca dell'Università di Milano sul fenomeno politico collegato all'uso di *twitter*, hanno evidenziato pari e a volte contrarie (politicamente) distorsioni rispetto alle indagini CATI tradizionali (www.youtrend.it, 2013, voicesfromtheblogs.com, 2013). Gli individui che si riconoscono nei gruppi politici più radicali, quelli di estrema sinistra o destra e gli stessi "grillini", sono stimati in misura minore rispetto alla loro reale consistenza che trova espressione nel voto politico⁶. D'altra parte, chi ha memoria e concreta saggezza circa la necessità di applicare forme probabilistiche nella selezione degli individui sa che è quasi inevitabile avere distorsione quando l'insieme delle persone monitorate è il frutto di un processo auto selettivo, tanto più nel caso specifico della rete, sottoinsieme del tutto parziale rispetto all'universo a cui si vorrebbero generalizzare i risultati ottenuti⁷. Interrogare la rete *sic et simpliciter*

⁶ Nell'analisi dei *twitter*, pur tentando di distinguere le affermazioni di quanti manifestavano simpatia verso un'area politica o un partito rispetto a coloro che facevano presumere una reale intenzione di voto, alcune liste politiche "radicali" hanno concretizzato un risultato minore rispetto alle attese mentre altre, come il Movimento 5 Stelle, è riuscito successivamente a coalizzare il voto di protesta in maniera superiore rispetto alle espressioni/*twit* ad esso favorevoli. In sostanza, sembrerebbe che la rete evidenzi il voto di protesta (forse) in misura maggiore rispetto i sondaggi più tradizionali, ma nel contempo non risolve le tipiche problematiche relative alla sovrastima (o sottostima) delle posizioni personali legate alla presunzione di "accettazione sociale" delle stesse.

⁷ Il 9 luglio 2013 il Presidente dell'AGCOM (Autorità di Garanzia nelle Comunicazioni) Angelo Maria Cardani, durante la presentazione al Parlamento della Relazione annuale ha sottolineato che il 37,2% degli italiani non si è mai collegato alla rete internet (media europea 22,4%). Il Presidente ha anche ricordato che "le famiglie che al 2012 avevano una connessione a banda larga su cavo erano il 49%, ma il dato di quelle connesse con almeno un minore balza al 71%" (AGCOM, 2013).

sarebbe come voler stimare il voto politico di una nazione in base ai partecipanti ad una manifestazione di piazza. Né gli *spare-panel* (consapevoli o inconsapevoli del loro ruolo, trattati come campioni probabilistici selezionati per un scopo di ricerca specifico), né alcune “nuove” modalità d’indagine rispondono appieno alla necessità di velocizzare i processi di raccolta dati, di renderli economicamente adeguati alla valenza dell’informazione acquisita e, nel contempo, di rendere statisticamente rappresentativo il campione e, con esso, credibile la ricerca sociale.

Come per ogni rivoluzione, tuttavia, è irrealistico immaginare che la restaurazione sia l’obiettivo da perseguire, come se nulla fosse accaduto o nulla di ciò che accade sia sostanziale. Come scrive Antonio Rao (2008, pag. 25) circa il modo corretto di affrontare la ricerca in economia «[...] il vero rigore, per gli studi economici, non consiste nell’inseguire l’irraggiungibile modello della fisica; consiste piuttosto, prima di tutto, nell’elaborare “categorie”, corrispondenti alla natura dei problemi che l’economia deve affrontare». Inoltre, Rao sviluppa l’idea di portare nell’alveo delle discipline sociali anche gli studi economici. In questo senso, quindi, se nella frase citata si sostituisce il termine “economici” con il termine “sociali/società”, risulta più chiaro il compito di chi si occupa di studiare i fenomeni in tale contesto. La critica, quale espressione di un atteggiamento libero da preconcetti, rappresenta certamente una delle categorie di cui parla Rao; ne costituisce un necessario punto di partenza. Meno lo è, quando essa assume un carattere specificatamente censorio, a volte fine a sè stesso, ovvero incapace non solo di paventare nuove proposte e confronti (favorire cioè il cammino della cultura della ricerca), ma incapace anche di cogliere la valenza distruttiva di alcune delle risposte fornite ai rilevanti problemi che si riscontrano nell’esercizio della ricerca sociale. Un atteggiamento precludente le nuove esperienze, quindi, non è il modo corretto di affrontare la situazione, così come, al contrario, è essenziale denunciare la pericolosità di alcune di queste scappatoie metodologiche. Per

comprendere meglio questa posizione, è forse utile ricordare il pensiero di Bertrand De Jouvenel (1964) che coniò l'espressione "idee sociali" per distinguerle dalla "cultura" concepita invece come percorso di selezione cosciente e regolato di una parte delle "idee sociali" presenti in una società. Egli sosteneva che le idee sociali si propagano all'interno della società attraverso il meccanismo della selezione naturale: l'una tentando di cannibalizzare l'altra. Come dargli torto? Come non pensare che l'anarchia tipica di questa rivoluzione tecnologica altro non sia che l'espressione della trasformazione stessa, dove è riscontrabile il fatto che tutti "fanno di tutto" e "tutto" sembra legittimarsi in quanto veloce e/o tecnicamente possibile? Il tentativo di molti istituti di legittimare l'uso degli *spare-panel* e di altri strumenti di ricerca, di prassi e di metodi d'indagine non probabilistici, altro non è che la propagazione naturale ed incontrollata di idee che, prima di tutto e forse anche al di fuori della sfera del desiderato, potrebbe avere come risultato l'abbattimento dell'attuale cultura della ricerca campionaria; ingombrante, sì, ma mai messa in discussione da teorie con altrettante solide basi scientifiche. Non è del tutto errato, quindi, affermare che l'assenza di un vero confronto con le opportunità offerte dalle nuove tecnologie comunicative tende a distruggere innanzitutto la cultura della ricerca scientifica sociale. Non cogliere questa sfida tecnologica così invasiva e non tentare di costruire un ponte tra le pratiche di ricerca e la teoria a cui fanno capo, significa adagiarsi su una delle due rive e, semplicemente, stare ad osservare quell'inevitabile fiume di critiche reciproche. L'opportunità di svolgere anche in Italia l'indagine scientifica internazionale ISSP (International Social Survey Programme) rappresenta un'occasione di rinnovare, innovare e far discutere di metodi, punti di vista e pratiche che sono di fatto l'anima stessa della ricerca sociale applicata. È questo il filo conduttore del mio lavoro di tesi: attraverso la descrizione di come si è implementato il disegno di ricerca ISSP Italia (per gli anni 2011-2012), cercare di analizzare pratiche, scenari e opportunità con cui la

ricerca sociale si sta confrontando o si dovrà necessariamente confrontare in ragione dell'avanzamento tecnologico.

Nella prima parte si affronteranno alcuni problemi inerenti i piani di campionamento e si esamineranno i cambiamenti intervenuti, nonché possibili, nelle tecniche e negli strumenti di somministrazione. L'intento sarà quello di comprendere come, quanto e soprattutto quali trasformazioni la tecnologia può indurre sulla cultura della ricerca empirica. Nello specifico, si cercherà di dare vigore all'ipotesi che la tecnologia, con il suo carattere inglobante e da molti considerato decisamente invadente, muta gli ambienti sociali, ibrida e trasforma gli strumenti di ricerca, permea i processi produttivi dell'informazione, e giunge, come già accennato, a predisporre le condizioni per un assalto alle basi stesse della cultura della ricerca applicata. Una particolare attenzione sarà posta ai processi di raccolta dei dati e alla conversione di quest'ultimi in informazione. Le trasformazioni tecnologiche, infatti, prendono vigore soprattutto in ragione della loro valenza economica, ma non sono quasi mai indifferenti rispetto alla qualità del prodotto di ricerca (l'informazione) poiché trasformano i processi di acquisizione del dato (in positivo o in negativo). La specializzazione produttiva, nello specifico, non è sinonimo di industrializzazione. La prima è funzionale all'abbattimento dei costi, la seconda attinge a queste economie e si pone lo scopo di riprodurre esattamente (se non migliorare) il prodotto finale. In questo senso, la metodologia della ricerca non può distogliere lo sguardo dall'assetto del comparto produttivo e presentarsi impreparata nel controllo di qualità dei propri processi, soprattutto in un momento in cui sono possibili pratiche così diverse tra loro. Il sistema produttivo generale, ad esempio, ha tentato di risolvere il controllo della qualità in produzione attraverso le certificazioni ISO. La metodologia della ricerca, dal canto suo, ci insegna in modo abbastanza preciso quali sono le fonti di distorsione a seconda della tipologia d'indagine. Ma è altrettanto attenta a riprodurre norme (per la certificazione o nei propri bandi di gara) a tutela dei processi? Uno sguardo

alla natura dei costi e a come i sistemi produttivi si evolvono, è sembrato quindi il modo opportuno per completare un primo quadro d'insieme. A conclusione della prima parte viene proposto un metodo di comparazione tra le diverse tecniche di ricerca: le curve di "costo metodologico". L'adozione di un indice di costo unitario, formalizzato in base alla copertura campionaria raggiunta, consente di evidenziare alcune peculiarità tipiche delle diverse tecniche d'indagine. Ad esempio, i sondaggi di opinione svolti tramite il CATI hanno notoriamente dei costi di contatto limitati, ma è altrettanto vero che il loro tasso di successo è generalmente minore di altre pratiche. Nello specifico, in ragione dei momenti della giornata in cui si svolge l'indagine, del numero di tentativi sviluppati per singola unità campionaria e della loro ripartizione nel periodo d'indagine, nonché in funzione del tempo complessivo concesso alla fase di raccolta dei dati, si possono raggiungere tassi di adesione differenti. Viceversa, le indagini *face to face* presentano costi fissi elevati, ma i loro tassi di successo sono notoriamente più elevati. Se partiamo dal presupposto che il grado di copertura campionaria raggiunto è uno dei principali indicatori di qualità, allora è opportuno rapportarlo al suo costo. In questo modo si possono paragonare tecniche diverse o è possibile stabilire fino a che punto è opportuno perseguire una determinata attività (di sollecito alla compilazione del questionario, ad esempio). Il costo metodologico, in sintesi, diventa uno strumento concreto attraverso il quale affrontare il tema dell'analisi dei processi di produzione del dato. In fondo, questo è lo stesso principio che si applica quando si deve determinare la numerosità campionaria. Tentare di azzerare l'errore di stima significa ad un certo punto innalzare esponenzialmente i costi, così come una scarsa numerosità determina un intervallo di confidenza troppo ampio per giudicare la statistica ottenuta un risultato attendibile. Valutare la curva degli errori campionari connessi alla numerosità significa – fissato un obiettivo di garanzia sulle stime ottenute – ricavare un costo, ossia un obiettivo numerico di campionamento.

La seconda parte introduce l'organizzazione ISSP, i temi affrontati e i vincoli metodologici imposti a tutti i 48 membri/nazioni che vi partecipano. Questo breve passaggio all'interno della tesi è strettamente funzionale allo sviluppo della terza parte, ossia la comprensione del percorso culturale-metodologico che il mondo scientifico internazionale, in particolar modo quello americano, ha elaborato per le proprie attività di ricerca sociale applicata.

Cercando di evidenziare gli elementi più caratterizzanti il contesto italiano, nella terza parte si chiariscono i motivi della crisi del CATI, il suo progressivo abbandono per le attività di ricerca scientifica e, soprattutto, il motivo per il quale – negli Stati Uniti in particolare – le indagini postali stanno vivendo una “nuova primavera”. La tecnologia, infatti, è entrata prepotentemente nell'ambito della logistica consentendo l'accumulo di conoscenze che in passato era praticamente impossibile rendere fruibili. Da qui nasce ciò che viene definita la rivoluzione geoinformativa, ovvero la possibilità di trasportare e analizzare sul piano territoriale i dati e le dinamiche inerenti gli elementi naturali e antropici (strade, edifici, flussi e/o concentrazioni produttive, di consumo, migratorie, ecc.). In questo panorama, sulla base di quanto studiato, applicato e visto in vent'anni di esperienza professionale nel campo della ricerca sociale, si concretizza lo sviluppo del piano operativo e di campionamento da me proposto per l'indagine ISSP. Esperienza del tutto innovativa, non solo per l'Italia ma anche nel contesto internazionale. Il piano di campionamento è stato progettato nel 2010 specificamente per le indagini ISSP in Italia e la sua applicazione inizia nel 2011 con l'indagine sulle disuguaglianze sociali. Negli Stati Uniti, all'incirca nello stesso periodo (2010), un gruppo di ricercatori del Dipartimento di Geografia dell'Università dell'Iowa formula per il NORC (*University of Chicago*) una proposta di campionamento di fatto analoga a quella italiana, destinata a un'indagine pilota della *General Social*

Survey nell'area metropolitana di Chicago⁸ (Kumar, Chen e Liang, 2011). Nel dicembre 2012, lo stesso gruppo di ricerca approfondisce la base teorica del campionamento areale e suggerisce le linee di sviluppo a livello nazionale del campionamento applicato a Chicago (Kumar, Liang e Linderman, 2012). Attraverso queste pubblicazioni, pur con le dovute differenze riferibili alle basi informative iniziali, si evince la sovrapposibilità delle due pratiche di campionamento (italiana e americana). L'Italia, quindi, sembrerebbe essere il primo paese in cui tale strategia campionaria è stata applicata su base nazionale.

La quarta ed ultima parte della tesi evidenzia alcuni aspetti organizzativi e i risultati del lavoro sul campo dell'indagine ISSP Italia. Lo scopo è quello di evidenziare eventuali elementi migliorativi, valutare l'efficacia delle geotipologie utilizzate per la stratificazione campionaria e, non ultimo, verificare se il CAWI (quale tecnica di autosomministrazione del questionario alternativa al postale) sia stata una scelta operativa adeguata.

⁸ La *General Social Survey* fornisce a ISSP i dati relativi agli USA.

1 – Riflessioni e Rifrazioni metodologiche

1.1 L'impatto della tecnologia nei disegni d'indagine

Volendo cominciare ad avvolgere quel filo che lega lo sviluppo tecnologico e la ricerca sociale è opportuno accennare ad almeno due aspetti: il primo è collegato alla possibilità che lo sviluppo tecnologico, grazie alla possibile convergenza dei vari strumenti di comunicazione, possa rappresentare una soluzione al problema della copertura campionaria; il secondo fa riferimento all'espansione delle opportunità di somministrazione del questionario e di controllo sulla qualità dei dati e, quindi, al cambiamento nelle prassi di lavoro che la tecnologia può permettere.

1.1.1 Lo scenario tecnologico e i problemi inerenti il campionamento probabilistico

1.1.1.1 Premessa

Come già detto in precedenza, la comunicazione, grazie all'evoluzione digitale, è diventata sempre più personale/individuale, spesso ridondante di mezzi e multiforme per modalità di contatto. Lo sviluppo tecnologico ha avuto come primo effetto quello di espandere enormemente le interazioni tra le persone, sia in termini di quantità di contatti che in termini di tempo dedicato⁹. La capacità di rimanere reciprocamente in contatto è talmente accresciuta che si potrebbe quasi affermare che è scomparso ogni potenziale isolamento; ovunque ci si trovi. Dal punto di vista della ricerca sociale, tale situazione sembra aver portato più scompensi che benefici. Paradossalmente, entrare in contatto con tutti i gruppi sociali, ossia ottenere un elevato grado di copertura dell'universo di riferimento, è diventata un'impresa più difficile di prima; più difficile di quando il telefono fisso era uno strumento distribuito pressoché uniformemente tra le famiglie (Istat, anni vari). A differenza di quel che è accaduto con il telefono

⁹ Nel 2012 gli italiani in media hanno giornalmente: trascorso 23 minuti sui social network e consumato servizi digitali per un totale di 4 ore e mezza; effettuato 10 minuti di chiamate telefoniche; mandato 21 mail e 30 tra sms e messaggi istantanei. Il consumo medio di televisione è stato invece di 115 minuti (AGCOM, 2013).

fisso, la cui diffusione ha alimentato anche quella informativa (si incrementava di giorno in giorno la consistenza dell'elenco telefonico), oggi si è aperto un divario tra lo sviluppo/diffusione delle tecnologie di comunicazione e le informazioni su coloro che le adottano. Internet e la telefonia cellulare, nello specifico, non hanno fornito pari conoscenze sulla distribuzione territoriale della popolazione e la sua raggiungibilità. Si è velocemente passati da un modello di relazione uno a uno, ogni tecnologia fisica (il telefono fisso, ad esempio) aveva una corrispondente informazione fisica (l'utenza telefonica - la famiglia), ad un modello molti a molti dove vari sono gli strumenti di comunicazione, così come molteplici sono i "volti" (*nickname*) e i ruoli con i quali le persone hanno la possibilità di presentarsi agli altri ed essere raggiunte. In tale situazione, sembra adatto parafrasare Pirandello affermando che la tecnologia ha assolto l'esigenza umana di potersi proporre, di volta in volta, come "uno, nessuno, centomila". In realtà, il problema della copertura campionaria non sta solo nella diminuita possibilità di contattare un soggetto (modalità e strumenti più idonei), ma anche nel farlo in modo indipendente dalla sua volontà. Infatti, ogni ricerca campionaria prevede l'applicazione di un'azione "forzosa" nei confronti dell'individuo selezionato (l'estrazione dall'insieme cui appartiene). Quando si parla di persone e non di cose, ciò non è poi del tutto indifferente. Di fatto, la disomogenea propagazione delle tecnologia tra la popolazione ha accresciuto notevolmente l'ansia del ricercatore contemporaneo. La sua attenzione è ora orientata, da un lato a comprendere quale scenario è ragionevolmente destinato a concretizzarsi in termini tecnologici e, dall'altro, quale mix di strumenti (telefono fisso, cellulare, web, posta tradizionale, ecc.) è opportuno impiegare. Il tema della copertura campionaria, d'altra parte, è strettamente collegato a quello delle possibili distorsioni di stima. In gioco c'è la credibilità e affidabilità stessa della ricerca applicata.

1.1.1.2 La convergenza degli strumenti di comunicazione

Prima di proporre degli argomenti sui quali focalizzare l'attenzione, è forse utile spendere qualche ipotesi e considerazione sull'evoluzione dei mezzi di comunicazione.

La tendenza futura sarà sicuramente quella di offrire pochi se non un unico strumento di comunicazione. Il fenomeno della convergenza strumentale, d'altra parte, è espressione diretta della praticità con la quale si deve necessariamente fare i conti. *Smartphone* e *tablet* sono chiari esempi del tentativo di fondere i mezzi di comunicazione (e non solo) con gli strumenti tipici dei pc. Sarebbe infatti paradossale se, per entrare in contatto con altre persone, negli anni a venire si dovessero usare più strumenti di quelli usati in passato: il telefono fisso, il cellulare, il pc, il fax, la posta tradizionale, quella elettronica ed altro ancora. La dotazione che si ha comunemente a disposizione all'interno di un ufficio, tenderà necessariamente a concentrarsi in un unico o in pochi strumenti, pur mantenendo le peculiarità tipiche dei diversi mezzi usati in precedenza. Ciò che si evolve, nella sostanza, sono le opportunità di ibridare le modalità di comunicazione. Ciò che si incrementa, è la possibilità di avere sempre tutto a portata di mano (*cloud computing*): i propri documenti, il proprio numero telefonico, la casella di posta elettronica, l'elaborazione dei dati, delle immagini, ecc.

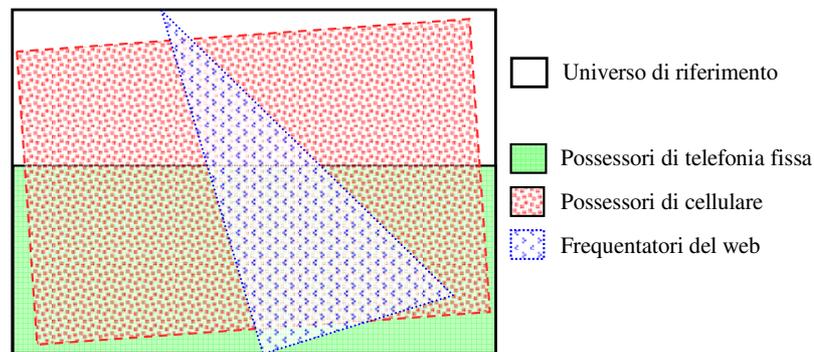
1.1.1.3 Il campionamento probabilistico: vittima del digital divide?

In attesa che la convergenza strumentale abbia il suo corso, nell'immediato rimane il problema di come affrontare la copertura dell'universo di riferimento in fase di realizzazione dell'indagine (raccolta dati).

È un'idea alquanto diffusa che la differente propagazione delle nuove tecnologie di comunicazione presso i diversi gruppi sociali (sub-universi tecnologici) altro non sia che una delle espressioni delle peculiarità socio-culturali ed economiche dei gruppi stessi (Sartori, 2006). Dal punto di vista

statistico, questi sub-universi tecnologici rappresentano quindi delle potenziali fonti di varianza per numerosi fenomeni oggetto di studio. Parrebbe opportuno, di conseguenza, far coesistere nella medesima indagine diverse tecniche di raccolta dati (il CATI ed il CAWI principalmente), auspicando così di intercettare la massima variabilità dei fenomeni indagati attraverso il coinvolgimento dei vari sub-universi tecnologici¹⁰. Ma è veramente così? Lo strumento di contatto o, per ora, un loro mix rappresenta oggi la problematica principale da risolvere nella ricerca campionaria? Per entrare nel merito della questione, è forse opportuno partire dall'analisi della situazione odierna.

Figura 1.1: Ricostruzione ipotetica dell'universo di riferimento per aree di sovrapposizione tecnologica



I frequentatori del web, i possessori di cellulari e quanti hanno ancora il telefono fisso in casa, possono essere trattati come universi distinti e, in quanto tali, è possibile somministrare loro la medesima indagine campionaria. Tali sub-universi, tuttavia, non sono sempre perfettamente distinguibili tra loro nello spazio di campionamento e durante il tempo dell'indagine. Si presentano come spazi sovrapposti e mutevoli: il possessore del telefono fisso ha quasi sempre anche un cellulare; l'assiduo frequentatore del web è quasi certamente connesso ad una linea internet veloce che, almeno per ora, è molto spesso collegata ad una linea di telefonia fissa; un *habitué* del web in ufficio potrebbe anche non essere un utente internet presso la propria abitazione. Che strategia utilizzare per sopperire all'impossibilità di coprire adeguatamente tutta la popolazione,

¹⁰ Accennando a questo argomento, Couper (2013) usa il termine *belief* (credenza/convinzione) aggiungendoci, immediatamente e tra parentesi, il termine *hope* (speranza).

tenendo conto della sovrapposizione dei sub-universi? Una lista dei cellulari non esiste e, in Italia, fare sondaggi (o ricerche in genere) con questo strumento è pressoché impedito a livello normativo (cfr. par. 3.1.1). Se, viceversa, si usa la sola tecnica CATI tradizionale, si escludono quanti hanno abbandonato la telefonia fissa. Ancora, se mai si trovasse un luogo virtuale all'interno del quale intercettare tutti i frequentatori del web, si escluderebbero quanti non posseggono un mezzo appropriato (pc, *smartphone*, ecc.). Con tali premesse, risulta evidente che sviluppare un piano di ricerca sperimentale che riesca a cogliere l'intero universo di riferimento attraverso l'adozione di diversi strumenti d'indagine (strettamente collegati ai sub-universi tecnologici), rappresenta un'attività che, se dal punto di vista tecnico-strumentale può rivelarsi piuttosto banale, non lo è altrettanto sotto l'aspetto metodologico. In altre parole, superate le problematiche strumentali e normative, rimangono aperte altre e non banali questioni. Citiamone un paio.

Quando alla mutevolezza dei contorni si somma la sovrapposizione dei sub-universi, le problematiche campionarie si acuiscono. Utilizzando più strumenti di contatto, risulta difficile eliminare sia la ridondanza informativa dettata dalla sovrapposibilità degli spazi tecnologici, sia la potenziale distorsione collegata alla parziale copertura dello spazio di riferimento. In sostanza, diventa arduo predisporre un piano di campionamento ottimale/efficiente e non distorto. D'altra parte, la tecnologia non sta stratificando l'universo di riferimento, ne sta semplicemente confondendo i contorni (sovrapposizione tecnologica). Dal punto di vista pratico ci si chiede quale probabilità di estrazione è di fatto attribuita all'unità campionaria che si trova in uno spazio di sovrapposizione. Se il piano di campionamento parte dall'utilizzo di una determinata tecnologia, avendo l'unità statistica diverse opportunità

d'essere contattata (o di rispondere¹¹), presenta una probabilità maggiore di essere estratta/rappresentata rispetto a coloro, ad esempio, che hanno abbandonato la telefonia fissa e che, senza avere a disposizione internet, usano semplicemente un telefono cellulare.

Un secondo elemento da considerare, ma che accenniamo solamente, è che l'impiego di diversi strumenti e tecniche d'indagine può incrementare la varianza estranea al fenomeno indagato amplificando l'effetto degli errori sistematici (Saris e Gallhofer, 2007). La tendenza all'unificazione degli strumenti di comunicazione molto probabilmente darà una svolta al problema della sovrapposizione dei sub-universi tecnologici. L'indagine che verrà condotta con quel futuribile, indeterminato, ma unico strumento tecnologico, permetterà di cogliere in modo più adeguato la varianza del fenomeno indagato. Tuttavia, se questo dovesse risentire della stessa problematica del web, ovvero determinasse un universo campionario sfuggente perché poco intercettabile e/o poco permeabile, allora l'insieme a cui si farà riferimento rimarrebbe comunque difficilmente aggredibile nella totalità dei suoi contorni. La necessità di riconoscere prima di tutto la "topografia" dell'universo rimane il problema da risolvere; una topografia che gli strumenti di comunicazioni e le tecniche d'indagine ad essi collegati, non ci aiutano più a conoscere come al tempo dello sviluppo della telefonia analogica/tradizionale.

La trasformazione tecnologica, in sintesi, potrebbe facilitare una soluzione al problema della copertura campionaria oppure costituire essa stessa un amplificatore delle difficoltà che già si incontrano oggi e che, in parte, erano presenti ieri. Nel periodo in cui la copertura offerta dalla telefonia fissa era pressoché totale, un analogo problema di copertura, la resistenza all'indagine di determinati gruppi sociali, rappresentava e rappresenta un macigno per la veridicità delle stime. È utile ricordare che

¹¹ Couper (2013) afferma che chi sta usando la nuova tecnologia (*web mobile*), spesso i giovani sempre così troppo assenti nelle indagini svolte presso la popolazione, hanno altrettante opportunità di rispondere anche attraverso i canali tradizionali.

la tecnologia solo in parte ha che fare con i veri problemi della copertura campionaria, i quali continuano ad essere:

- la possibilità di identificare il contorno esatto dei sub-universi (o dell'universo) di campionamento (la reale presenza di una lista di riferimento idonea agli scopi campionari);
- l'accessibilità dello spazio di campionamento (accessibilità alla lista e accessibilità all'unità di campionamento)
- la disponibilità dell'oggetto di osservazione (le mancate risposte).

Al fine di focalizzare l'attenzione su questi temi, due sono gli esempi che potremmo citare per asserire che il problema dell'identificazione dell'universo è disgiunto dalla copertura offerta dalla tecnologia impiegata per rilevare i dati.

1. Il cellulare e il web sono strumenti concepiti per essere di fatto indifferenti rispetto al territorio. Quand'anche tutta la popolazione ne fosse provvista, i tipici campionamenti su aree territoriali limitate (un comune, una provincia, ecc.) non sarebbero facili da concretizzare. L'assenza di una lista riferibile a specifiche ripartizioni dell'universo ingenera problematiche complesse e indipendenti dalla copertura complessiva offerta dalla tecnologia attraverso la quale possono rispondere le unità contattate.
2. Le incognite collegate ad uno scenario (universo) mutevole sono analoghe a quelle che si possono incontrare utilizzando una lista vetusta. Se gli aggiornamenti non sono adeguati, una lista potrebbe riportare la presenza di persone oramai non pertinenti per l'indagine, ma soprattutto denotare l'assenza di altrettante figure che nel frattempo sono entrate a pieno titolo nell'universo di riferimento. Di conseguenza, tanto più veloce è il mutamento, tanto maggiore sarà il rischio di coprire l'universo in modo parziale. Durante l'espansione della telefonia fissa, questo problema è stato risolto attraverso la generazione casuale di numeri telefonici (RDD- Random Digit Dialing); pratica messa in atto in

base a specifiche e concrete conoscenze su come i sistemi telefonici si evolvevano, ossia su come venivano generati i numeri telefonici che erano assegnati alle utenze. Con la tecnologia digitale, tali strategie di individuazione sperimentale dei contorni dell'universo non sono di fatto applicabili.

La velocità dei cambiamenti dell'universo d'indagine e la necessità di definirne il contorno esatto è quindi una questione indipendente dalla tecnica di contatto. A volte la soluzione al problema si è concretizzata con una tattica che ha visto la stessa tecnologia come protagonista (l'RDD), ma ciò non scalfisce il fatto che quella della copertura rimane una questione più ampia e, di fatto, indipendente dagli strumenti di contatto.

Per quanto riguarda l'accessibilità della lista, dobbiamo prendere atto che le critiche rivolte al *cloud computing*, cioè l'essere una potenziale trappola di marketing (R. Stallman, 2008), valgono anche per le questioni relative al campionamento. Se la raggiungibilità dei soggetti dovesse passare tramite le comunità virtuali, verrebbero seriamente minate le basi dei campionamenti probabilistici; sia che si osservi (banalmente) la questione dal punto di vista dell'accessibilità della lista, sia che la si osservi dal punto di vista della sovrapposizione e attendibilità strutturale dei sub-universi di riferimento. Nulla vieterebbe la partecipazione di un soggetto a molteplici e potenzialmente sconosciuti *social network*, così come nulla garantirebbe la proliferazione di false identità presso le comunità virtuali¹². Appena tentiamo di ricomporre l'universo non più come insieme di sovrapposti sub-universi tecnologici, quando, piuttosto, quale diversità rispetto ai luoghi virtuali frequentati, l'immagine definita in precedenza non cambia. Le difficoltà di operare con un adeguato piano di campionamento, di fatto, rimangono immutate.

¹² A tale proposito si rimanda a quanto recentemente pubblicato da M.P. Couper (2013) circa il rischio di manipolazione dei dati relativi alle comunità virtuali. Vengono citate, ad esempio, le cospicue attività di propaganda condotte da organizzazioni spionistiche militari; l'8,7% di false identità stimate tra gli utenti di Facebook; il 50% di false identità tra i 37 milioni di *twitter follower* di un famoso cantante americano, ecc.

A partire da queste considerazioni, potremmo affermare che la propagazione dei fenomeni in un ambiente mutevole, spesso indistinto e/o poco permeabile è, per la ricerca sociale, un impegno teorico e pratico che rimarrà sempre e comunque consistente e pressoché indipendente dalla tecnologia in uso. Al più, dato il cambiamento tecnologico e le conseguenze che ne derivano, sarà gravato anche dalla problematica relativa all'accesso democratico all'informazione di base (la lista). Seguendo lo schema analitico proposto da Rao (2008), è quindi appropriato affermare che la copertura dell'universo costituisce una problematica solo in parte connessa al mutamento tecnologico contingente. Più propriamente, essa è collegata alle trasformazioni dell'ambiente sociale indotte dal cambiamento tecnologico; tra queste non si possono trascurare quelle relative alla cultura scientifica e alle idee sociali e quelle legate alla conoscenza di base dell'universo; conoscenza che sempre più si concentra nello spazio virtuale del web (fonti dati). Sui cambiamenti culturali imposti dalla tecnologia citeremo alcuni elementi di dibattito nel prossimo paragrafo. Sulla seconda trasformazione, ossia sulla concentrazione informativa nel web, ritorneremo a parlare nella terza parte della tesi.

Per concludere potremmo dire che una posizione metodologicamente corretta non nega la possibilità che il *digital divide* possa essere causa e nel contempo effetto di talune diversità osservabili nella popolazione. È necessario, tuttavia, rivedere le pratiche di campionamento alla luce dell'evoluzione tecnologica affinché non si diffonda una mistificazione culturale che tende ad innalzare il modo e gli strumenti utilizzati per la raccolta dei dati (quale soluzione delle problematiche di copertura campionaria) invece di soffermarsi a coniugare il metodo scientifico-campionario con la trasformazione in atto. A volte, direi, è necessario tornare ad interrogarci addirittura sulla semantica dei termini usati. Quando Couper (2013) afferma che chi risponde alle indagini web *mobile*¹³ potrebbe tranquillamente farlo anche con altri strumenti di comunicazione

¹³ intendendo con questo l'uso del CAWI su smartphone

più tradizionali, pone una questione non banale. La tecnologia, infatti, ha certamente il carattere d'essere una *comodity*¹⁴, sia per l'utente-intervistato sia per il ricercatore. Su tale forza, essa fonda normalmente la sua capacità di diffondersi. Al momento del lancio sul mercato, un oggetto tecnologico può pure diventare appetibile commercialmente perché qualifica l'utente come *trendy* (anticipatore o, poi, aderente a una moda), assumendo quindi un carattere di *exclusive* (differente, ristretto a pochi, qualificante per l'appunto). È tuttavia necessario interrogarsi quando e quanto nella vita quotidiana la tecnologia sembra mutare il suo carattere d'essere una *comodity trendy/esclusive* diventando così un oggetto semplicemente *esclusive-divide*, ossia un ostacolo che separa nettamente gruppi di persone. Pur non volendo minimizzare il problema dell'alfabetizzazione digitale delle persone, sorge il dubbio che anche nel mondo della ricerca applicata si tenda a vendere, così come un oggetto di consumo qualsiasi, le proprie idee e convinzioni (*belif*, diceva Couper, 2013) eliminando qua e là qualche parola (*comodity*) ed esaltandone, come fossero i punti di forza del prodotto in vendita, altrettanti concetti, parole, punti di vista: passando così da *trendy, exclusive a divide*. Per tale motivo, come proposto da Rao (2008) citando l'opera di Schumpeter (1962) dedicata a "scienza e ideologia", è opportuno palesare anche le posizioni di valore che si assumono quando si discute e ci si interroga sul ruolo, in questo caso, degli strumenti di comunicazione nella ricerca sociale. Senza un punto di vista, affermava Schumpeter, non si ha una visione, perché: «Sebbene si proceda lentamente, a causa delle nostre ideologie, non si potrebbe camminare affatto senza di esse» (Schumpeter, 1962. pag. 276).

1.1.2 Interazioni tra tecnologia, cultura di ricerca e idee sociali

Cogliere la valenza delle trasformazioni culturali nel momento in cui queste si manifestano è un'impresa azzardata. Le idee mutano, a volte

¹⁴ Così come lo fu il telefono fisso all'epoca in cui questo strumento rappresentava una novità tecnica rispetto alla posta tradizionale o al contatto diretto (Groves, e Kahn. 1979).

molto velocemente, e quasi mai seguono un percorso lineare. Le trasformazioni culturali, al contrario, si sedimentano in tempi più o meno lunghi e necessitano, innanzitutto, di una sostanziale accettazione affinché possano propagarsi. Come scriveva John Kenneth Galbraith (1958, pag. 37): «Le idee finiscono per mobilitarsi in ragione di ciò che la comunità tutta o auditori particolari trovano accettabile». La compresenza di diversi punti di vista e posizioni contrastanti è una situazione connaturata alla trasformazione stessa. Una condizione, quest'ultima, caratterizzata da un'elevata variabilità quando più la realtà sociale muta repentinamente con la diffusione stessa (in base alla forza) delle nuove idee. Il ricercatore, così come l'uomo comune, agisce in base alle proprie visioni e convinzioni e ciò definisce lo spartiacque sostanziale tra le tematiche affrontate nelle scienze naturali e tutto quello che, al contrario, riguarda la società umana (Rao, 2008). Per comprendere quanto i cambiamenti tecnologici incidano sia sulla cultura di ricerca sia sulle idee sociali, è forse utile citare alcuni ambiti di confronto presenti tra i ricercatori. Primo tra tutti, il modo con cui è percepita, considerata e trattata la distorsione campionaria indotta, per l'appunto, dalla tecnologia.

1.1.2.1 Universi liquidi e riponderazioni delle stime

Seguendo un'impostazione analitica "classica", molti ricercatori focalizzano l'attenzione sulla copertura fornita dalle diverse tecnologie e, rispetto al tema d'indagine, formulano ipotesi (o acquisendo dati) su quale potrebbe essere il grado di difformità di ogni gruppo tecnologico indagato rispetto all'ipotetico valore medio (o di ciascun gruppo rispetto ad ogni altro). Con i due parametri ottenuti (copertura e difformità), pur usando un solo mezzo di contatto (il telefono fisso ad esempio), è potenzialmente possibile calcolare ed eliminare il grado di distorsione sui risultati d'indagine. Tanto più piccole sono le differenze che si trovano nella popolazione variando il mezzo di contatto, tanto minore sarà l'errore commesso (distorsione). In particolare, la sua entità sarà proporzionale alla quantità di soggetti che non si avvalgono della tecnologia utilizzata per il

contatto (Di Gioia 2009). Al contrario, se i differenziali sono elevati anche piccoli gruppi di popolazione non intercettati con un determinato ausilio tecnologico potrebbero incidere in modo rilevante nella stima finale. La consistenza dei gruppi di soggetti a cui si fa riferimento e le loro specificità, in sostanza, sono parametri che qualificano l'attendibilità dei risultati raggiunti¹⁵.

La riponderazione delle stime, pur essendo metodologicamente corretta, non è esente da critiche. Innanzitutto spesso non è tenuta in debita considerazione l'eventuale sovrapposizione dei sub-universi tecnologici e la loro struttura di interazione¹⁶; in secondo luogo gli stessi sub-universi sono fortemente dinamici. Dal momento che la riponderazione propaga gli errori in forma quadratica (Diana e Salvan, 1993), per applicarla è necessario avere un'adeguata quantità/qualità di informazioni per ogni sub-universo e per ogni carattere d'indagine¹⁷. In caso contrario, la pratica potrebbe rivelarsi disastrosa producendo stime fortemente variabili e ancor più imprecise (distorte). Qualora si optasse per la riponderazione, quindi, la procedura andrebbe metodologicamente approfondita e, soprattutto, resa palese in fase di diffusione dei risultati. Il costo economico per ottenere dei parametri utili alla riponderazione potrebbe essere così elevato da far preferire un'indagine condotta con i tradizionali sistemi (*face to face o postale*) piuttosto che rivolgersi alle tecnologie (indagini *mixed mode*). È questa la strada perseguita da vari ricercatori che preferiscono sistemi di contatto tradizionali (Kaplowitz et

¹⁵ Se il carattere indagato è trasversale (indipendente) rispetto alle sub-popolazioni definite dalla tecnologia usata per il contatto, allora la distorsione risulterà nulla per qualsiasi strumento di comunicazione utilizzato nella rilevazione dei dati.

¹⁶ La sovrapposizione dei sub-universi campionari impone una complessità di calcolo spesso ignorata. Quindi, per comodità si ipotizzano sub-universi disgiunti (Fuchs e Busse, 2009).

¹⁷ È possibile, ad esempio, procedere con dei campionamenti doppi acquisendo informazioni preliminari sui sub-universi d'indagine in modo da disporre di parametri di riferimento per effettuare successivamente adeguate ponderazioni.

al. 2004)¹⁸ sobbarcandosi costi per unità di contatto più elevati, anche se, come si vedrà meglio in seguito, non necessariamente maggiori in riferimento alla spesa totale una volta rapportata alla qualità del risultato (cfr. par. 1.3.1).

Contrapposta e ancora più radicale è la posizione di altri ricercatori i quali sostengono che la realtà attuale è frutto di una profonda trasformazione della società e che, in tale situazione, l'intero approccio metodologico all'indagine campionaria deve essere sostanzialmente rivisto¹⁹. Sintetizzando la loro posizione e volendo fare un paragone spesso usato nella didattica, si potrebbe dire che l'urna, dalla quale si pescano le palline bianche e quelle nere, si è improvvisamente dissolta, rendendo il contenuto del tutto "liquido". In tale situazione, pensare ad un universo distinguibile ed intercettabile (mediaticamente, ad esempio) è un approccio incoerente con la l'attuale realtà così dinamica e globale. Tale posizione, che mette insieme gli aspetti riferibili ai comportamenti umani con quelli tecnologici, porta questi ricercatori a favorire forme di studio non più basate sul campionamento e sulla casualità della selezione degli individui, quanto sull'analisi dei fenomeni presenti (e passati/accumulati) in rete. Studiare la dinamica della società attraverso il web è un punto di vista, più propriamente un'impostazione culturale, che trae forza dall'idea d'essere di fronte ad un processo di concentrazione informativa promosso dalla rivoluzione digitale e da internet in particolare. Tempi e modi dei cambiamenti della società sono congiuntamente presenti nel contenitore di

¹⁸ L'autore mette in evidenza come la classica strategia dell'invio postale del questionario cartaceo (composta da lettere di presentazione dell'iniziativa di ricerca e successivi solleciti alla compilazione), raggiunga il tasso di successo più elevato (31,5%). Un livello di adesione che, nel gruppo sperimentale "colpito" dalla Web-Survey, è quasi raggiunto solo se l'unità di campionamento è stata precedentemente avvertita con una lettera tradizionale (29,7%). Il solo invito via mail alla compilazione della web survey, al contrario, comportava un tasso di successo inferiore di 10 punti percentuali (da 31,5% a 20,7%). Ovviamente, tutte le unità campionarie erano utenti/possessori di internet. I costi delle varie strategie di contatto erano differenti, così come si sono accertate differenze significative nelle caratteristiche socio-demografiche tra i vari gruppi di rispondenti.

¹⁹ Di tale posizione ne sono esponenti primi quanti stanno promovendo fortemente l'uso degli *spare-panel* nella ricerca. Il riferimento principale è costituito dalle teorie di Bauman (riprese da Bordoni nel suo saggio del 2013).

internet. Il web rappresenta la telecamera con cui osservare la società e, in sostanza, il film che se ne può trarre (i *big data*²⁰ e il loro accumularsi nel tempo) potrebbe essere maggiormente esplicativo della fotografia scattata con un'indagine. Visto in questa ottica, prende maggiore forza il tentativo, citato in premessa, di prevedere il voto politico 2013 attraverso lo studio dei *twitter* lanciati dal "popolo del web", ovvero attraverso l'analisi di coloro che seguivano e commentavano le idee e i fatti inerenti i vari esponenti politici, i partiti e i gruppi di dibattito socio-politico in Italia. Gli errori commessi si sono rivelati paragonabili a quelli presenti nei dati forniti da altri istituti di ricerca che avevano affrontato lo stesso tema attraverso tecniche più tradizionali: indagini CATI su telefonia fissa, indagini miste CATI-CAWI, ecc.²¹. Lo studioso di metodologia della ricerca affermerebbe che tutte queste esperienze, sia quelle condotte sul web sia quelle promosse con tecnologie più tradizionali, sono affette da problemi di copertura campionaria, da distorsioni indotte dal grado di partecipazione, se non addirittura del tutto deformate dalla scarsa qualità degli *spare-panel* che spesso vengono utilizzati o assommati ai soggetti casualmente contattati (Di Gioia, 2009). In sostanza, tutte iniziative che lasciano consistenti dubbi circa l'impostazione probabilistica della ricerca.

1.1.2.2 Numerosità campionaria e omogeneità dei sub-universi tecnologici: il mito dei rispondenti fidelizzati

Come più volte detto, le trasformazioni indotte dal digitale portano inevitabilmente a una reazione sul piano pratico della ricerca applicata. Sostanziale, dal punto di vista metodologico, è verificare se per propagandare o per giustificare le proprie scelte, ci sia la tendenza ad attribuire a queste reazioni il carattere di obbligatorietà e/o necessità, mistificandone così gli elementi del dibattito. Si fa riferimento, in prima istanza e quale ulteriore esempio, alla propagazione degli *spare-panel* e

²⁰ I *big data* sono *dataset* eterogenei, riferibile a varie tipologie di dati, provenienti da fonti diverse (strutturate e non) (http://it.wikipedia.org/wiki/Big_data)

²¹ Circa le dissimilarità tra la consistenza dei fatti reali e la corrispondente discussione registrata sui social network (Twitter o Google Flu Trends), si veda quanto riportato da Couper (2013).

all'indifferenza con la quale vengono ibridati i campioni (sperimentali o meno che siano) con i panel (veri o presunti che siano²²). Ciò a fronte di una reale difficoltà ad ottenere un'adeguata copertura dei sub-universi tecnologici, cioè all'assenza di una lista di riferimento per i singoli canali mediatici utilizzati nella rilevazione dei dati.

Una prima scorrettezza nel dibattito fa riferimento alla numerosità campionaria. È patrimonio della cultura della ricerca applicata la convinzione che la sola numerosità non può essere definita come l'indicatore unico della bontà di un'indagine (Diana e Salvan, 1993), ovvero della capacità di fornire stime affette da limitati e controllati livelli d'errore. Eppure, sempre più spesso al concetto di metodo campionario (selezione probabilistica delle unità) è sostituita l'idea che una certa quantità di casi sia sufficiente per ottenere adeguate informazioni sul fenomeno d'interesse. In quest'ottica, ogni costo volto a garantire la casualità nella scelta dell'individuo appare superfluo e un campionamento non probabilistico o uno *spare-panel* assumono la stessa valenza di un campione sperimentale.

Successivamente si enfatizzano le diversità dei vari sub-universi tecnologici, così che il fine giustifichi i mezzi. Se è accettabile l'ipotesi che il grado di utilizzo e/o il possesso di particolari strumenti tecnologici sia anche espressione di talune differenze tra sub-universi, non lo è affatto la

²² Relativamente ai panel propriamente detti, si deve sempre avere presente che la tecnica di costruzione, la loro stabilità nel tempo e l'eventuale grado e modalità di aggiornamento delle unità che vi partecipano in tutto il periodo di utilizzo, rappresentano elementi sostanziali rispetto all'universo verso cui possono essere generalizzati i fenomeni sondati. Se non sono applicate delle prassi specifiche in relazione ai citati elementi, l'uso del panel è considerato adeguato solo per la stima degli effetti riscontrabili nel tempo in una determinata coorte, ossia adatto alle indagini longitudinali riferibili alla popolazione inizialmente coinvolta nel panel stesso. Di conseguenza, non possono essere confusi o spacciati quali tecnica alternativa ed altrettanto valida all'inferenza sulla popolazione generale così come – al netto di vari altri elementi discorsivi – è possibile fare con i campioni probabilistici.

pratica di utilizzare gli *spare-panel* quale emblema di un determinato sub-universo tecnologico, sia per le indagini via web sia per quelle telefoniche²³.

Il richiamo alla coerenza ai principi metodologici cardine della ricerca empirica consente di evitare degli errori dettati dall'ansia di cogliere la modernità senza interrogarsi sulla natura stessa delle trasformazioni indotte dalla tecnologia. Dal punto di vista statistico, avvallare una pratica che vede la scelta delle unità campionarie con un metodo privo di casualità e piegare l'indagine al solo fatto che si utilizzi un determinato mezzo tecnologico, significa sostenere una sostanziale omogeneità in termini di caratteristiche di coloro che hanno adottato un certo strumento di comunicazione. Il concetto di rappresentanza estesa²⁴, a volte utilizzato con oculatezza e metodo in alcune tecniche d'indagine su testimoni privilegiati (Delphi, Nominal Group Technique, ecc.) (VandeVen e Delbecq, 1974), non appartiene alla cultura scientifico-probabilistica del campionamento. L'imprudenza di tali pratiche sta nel fatto che nello stesso momento in cui si afferma la necessità di sondare i diversi sub-universi tecnologici con altrettanti e collegati strumenti tecnologici, si stabilisce in modo del tutto arbitrario una struttura di variabilità dell'universo di campionamento. Si impone che la varianza totale del fenomeno indagato sia espressione della diversità tecnologica (varianza tra gruppi) piuttosto che risiedere anche, o soprattutto, all'interno dei gruppi stessi. Questa ipotesi è inaccettabile da molti punti di vista: non solo perché risulta inverosimile per tutti i fenomeni indagati, non anche perché è evidente che i sub-universi tecnologici sono sovrapposti e mutevoli nel tempo, ma soprattutto perché nega la componente oggettiva del *digital divide*, il quale seleziona prima di tutto per territorio (disponibilità della connessione tecnologica) e solo successivamente per caratteristiche del singolo. Soprassedere al principio

²³ L'uso degli *spare-panel* è una pratica che coinvolge anche la tecnica CATI con telefonia fissa o mobile. La loro adozione ottempera alla necessità di entrare in contatto, in modo assai economico per tempi e costi, con gruppi di soggetti difficilmente raggiungibili quali i giovani, la popolazione attiva, ecc.

²⁴ Capacità di un singolo individuo di essere portavoce di un gruppo omogeneo.

di casualità quale elemento per la salvaguardia della copertura campionaria, cioè sostenere un'ipotetica omogeneità entro i sub-universi tecnologici è decisamente un azzardo se non un cospicuo, spesso volontario, errore. Anche perché, da un punto di vista più generale, si dovrebbe poi spiegare come l'adozione stessa di una determinata tecnologia di comunicazione ingeneri in chi l'adotta delle repentine omologazioni comportamentali, di atteggiamento se non anche valoriali. Può la tecnologia (come qualsiasi altro fenomeno di consumo di massa) essere un "trasformatore" ideologico²⁵? Se una determinata tecnologia di comunicazione dovesse essere pervasiva tanto quanto lo è stato la telefonia fissa, dovremmo aspettarci uno stato di omologazione diffuso? La storia, se non anche la logica, ci propone una situazione del tutto contraria: tanto più si diffonde un fenomeno (specie di consumo), tanto più decresce l'omogeneità interna al gruppo in questione. Ricordiamo, a tal proposito, quanto affermato da Bertrand de Jouvenel (1964) circa la distinzione tra idee sociali e cultura, quest'ultima definita quale insieme di idee sottoposte a selezione, dibattito e, bisognerebbe aggiungere nel caso della scienza, a confutazione. Se è ragionevole supporre almeno due piani di lettura delle idee, quello delle idee sociali e quello delle ideologie o della cultura, allora sembra difficile accettare l'ipotesi che la tecnologia possa omologare profondamente e celermente la popolazione. Semmai, parafrasando il pensiero di de Jouvenel, essa rappresenta l'arma con la quale le idee sociali tendono a cannibalizzarsi reciprocamente.

Potremmo concludere dicendo che se si riporta nei binari della cultura sperimentale sia il tema della numerosità campionaria (misurabilità dell'errore commesso in funzione della quantità dei casi sondati), sia il misticismo connesso alla presunta capacità dell'universo tecnologico nel dare spiegazione ai sempre diversi fenomeni sociali, si deve convenire che i

²⁵ A. O. Hirschman (1990, pag. 281) afferma che «[l'ideologia è] un (moderatamente coerente) insieme di idee e di credenze che si forma all'interno di una società, ne spiega i fenomeni socio-economici, e contiene altresì qualche norma per l'azione».

mezzi non sono tutti giustificabili per il fine. In particolare, diventa difficile sostenere che si possano evitare distorsioni di stima quando la selezione dei soggetti intervistati non è accompagnata da un criterio di casualità (e non certo di natura tecnica-strumentale) o da una scelta oculata dettata da confermate ipotesi circa la struttura di variabilità dell'universo di riferimento. D'altra parte, la presenza di elevati tassi di non risposta nelle indagini spingono a dire che nemmeno l'adesione casuale e spontanea tutela da potenziali distorsioni; a maggior ragione quella assodata (*spare-panel*), il mix di fonti e di tecniche e metodi di ricerca che vengono adottati non come "condizione ambientale" predisposta al fine di incrementare il tasso di adesione (Dillman et al. 2009), quanto - piuttosto - come indipendenti universi tecnologici dai quali attingere al fine di raggiungere una determinata numerosità "campionaria" (o struttura socio-demografica del *dataset* finale). Quest'ultima, sembra molto più una pratica svolta al prevalente scopo di rendere veloci, numericamente cospicue e poco costose le rilevazioni piuttosto che una oculata strategia che punta alla conoscenza dello stato di fatto. Potremmo concludere dicendo che, sdoganato metodologicamente il *mixed mode* (al netto degli effetti di *question-wording* - Bernardi e Tuzzi 2005) quale prassi di ricerca, questo viene travisato quale tecnica di campionamento; due ambiti che non devono di certo essere confusi.

1.1.2.3 La credibilità della ricerca sociale: in bilico tra cultura aristotelica e metodo scientifico

La presenza di così diversi punti di vista e la quasi indifferenza con la quale queste pratiche di ricerca vengono applicate (perché a loro volta comunemente accettate), fanno sorgere il dubbio che, così come accade per molti altri aspetti della vita sociale, anche nella ricerca applicata si stia ritornando ad una cultura aristotelica e pre-scientifica circa la credibilità delle informazioni (Gili, 2005). In sostanza, l'informazione ritorna ad assumere autorevolezza in base a chi ne è il promotore e non alla tecnica e prassi adottata per produrla. Il contrappeso metodologico sembra aver

perso nel tempo la sua valenza o, quantomeno e per quel che riguarda i produttori dei dati, sembra culturalmente dimenticato. Al più, si potrebbe dire, reso del tutto relativo e opinabile. Le sfide scientifiche partono anche dal tentativo di falsificare le teorie precedenti e, in tal senso, è plausibile e opportuno ogni tentativo. Rimane tuttavia il dubbio, come già affermato, che la propagazione di alcune posizioni non sia tanto frutto di un processo scientifico (una cultura scientifica), quanto la necessità di accreditare/giustificare le prassi adottate e di dare una risposta pratica alle trasformazioni tecnologiche e sociali. Ciò, ovviamente, ha delle ripercussioni nella società. La presenza di contrapposte informazioni, tutte teoricamente offerte con pari e indistinta credibilità, pone sullo stesso piano, quello dell'indifferenza e dello scetticismo, l'atteggiamento dei più. Sarebbe interessante valutare se il ricercatore sociale gode di maggiore o minore credibilità rispetto a chi si occupa di fisica, chimica e altre scienze naturali²⁶. Sarebbe interessante fare questo raffronto dopo che una platea di milioni di persone hanno assistito a trasmissioni televisive di approfondimento politico-sociale in cui si sono trovati di fronte due istituti di ricerca che hanno avvallano e presentano dati diversi. Dove lo scontro sfocia nell'enunciare i propri titoli, piuttosto che ammettere che i dati proposti, al di là della numerosità sottostante, presentano tali e tanti problemi di copertura dell'universo al punto da dover riponderare le stime in modo – direi – quasi del tutto soggettivo. Riponderazioni che producono intervalli di variazioni che coprono ampiamente le differenze tra i dati presentati dai due contendenti. La credibilità della ricerca sociale, ma anche economica e di altre scienze e discipline umanistiche, non è argomento di tesi. Rimane il fatto che un atteggiamento eccessivamente relativista tra coloro che praticano la ricerca socio-economica, capace di negare la relazione anche dei fatti più evidenti ed oggettivi, non può ritenersi

²⁶ «Di nessun'altra disciplina, ritengo, esiste una *summa* così corposa di concettualizzazioni sapienziali, a testimonianza di una diffidenza, forse di un timore, così spontanei e diffusi», scrive il prof. Lorenzo Bernardi (2001, pag 141) commentando il libro di J. Bibby (ed. 1983) che raccoglie «qualche centinaio di citazioni relative alla statistica, alla sua funzione, più spesso alle sue perfidie e perversità».

estraneo né al modo con il quale è affrontato il dibattito scientifico né a ciò che queste scienze sono in grado di produrre in termini di risultati. Elementi che inevitabilmente penetrano poi nelle idee sociali. La credibilità della ricerca sociale applicata, valutata alla luce del tentativo di superare le problematiche proposte dagli sviluppi tecnologici, è un argomento che ci porta lontano dall'obiettivo di analizzare le questioni di campionamento e le metodiche di rilevazione²⁷. Anch'essa, infatti, è da ritenersi causa e contemporaneo effetto delle trasformazioni. Per tale motivo si preferisce orientare l'analisi verso gli effetti concreti dei cambiamenti in atto come ad esempio è il tema della privacy.

1.1.2.4 Idee Sociali: la tecnologia e la privacy

Nell'ultimo decennio si è aperto un interessante fronte di dibattito riguardante l'impatto che può avere la tecnologia sulla libertà personale o, meglio, sul diritto di ognuno di noi di conservare un ambito di riservatezza. Le posizioni più estreme giudicano l'avanzamento della tecnologia quale spettro del nuovo millennio. D'altra parte, come negare l'interesse economico che ruota attorno alla possibilità di osservare l'individuo nel suo utilizzo del web? Necessità, *network* di amicizie, ansie, e altro ancora, possono essere intercettate in modo da promuovere un prodotto o un servizio al momento opportuno²⁸. Le quotazioni economiche stratosferiche attribuite ad alcuni siti internet (motori di ricerca) o quelle relative ad alcuni *social network*, ad esempio, vengono percepite da molti come valore connesso alla possibilità di osservare la persona, più che come valore dello spazio "fisico" pubblicitario inserito in un sito. In altre parole, la tecnologia diventa un moltiplicatore economico connesso alla possibilità di mettere in

²⁷ Su tale tema è interessante il libro di Raffaele De Mucci (1989) "*La forma scienza. Scienza e scienze sociali*", Cleup.

²⁸ A differenza di un tempo, oggi le pubblicità in internet sono spesso "territoriali" (legate al luogo da cui si accede alla rete) e/o distinte in ragione di determinate caratteristiche possedute da colui che sta navigando. L'analisi dei profili degli utenti internet ha fatto sorgere nuovi termini e nuove pratiche di ricerca come il *web-mining* (applicazione di tecniche d'analisi di *data-mining* specificatamente pensate per il web).

relazione la pubblicità di un prodotto o di un servizio con i momenti in cui l'esigenza è espressa.

Al di là di questi scenari, che si riveleranno più o meno proficui dal punto di vista economico, rimane il fatto che l'uso massiccio di questi nuovi strumenti ha contratto decisamente la disponibilità del cittadino a partecipare alle indagini. È questa l'opinione di molti ricercatori circa i motivi di una crisi che sta vivendo la ricerca in generale ed il CATI in particolare (Battaglia e altri 2008). Le continue telefonate a scopo commerciale, quelle per sondare l'opinione dei cittadini, l'utilizzo di sistemi di telefonia in grado di generare più chiamate rispetto alle reali possibilità di gestione con gli intervistatori attivi in quel momento (fenomeno dello *spam* telefonico e via posta elettronica), ha costretto i vari governi a declinare risposte legislative a tutela dei cittadini e a scapito dell'opportunità di entrare in contatto con gli individui (vedasi Registro delle Opposizioni; normative sull'uso del *predictive dialing*; tracciabilità telefonica dell'istituto di ricerca, ecc. presenti nel paragrafo 3.1.1). Ciò ha delle pesanti ripercussioni dal punto di vista sia metodologico (la possibilità potenziale di entrare in contatto con l'intero universo) che delle idee presenti nella società, da cui dipendono a cascata le decisioni legislative. La disponibilità dell'oggetto di osservazione è sempre più precaria, così come lo è il distinguo (legislativo e di percezione tra la popolazione) tra una scienza rivolta al governo dell'evoluzione sociale rispetto a quella connessa all'ambito strettamente economico-commerciale. L'abbattimento dei costi prodotto dallo sviluppo tecnologico, ha portato ad una esplosione delle iniziative di ricerca. Ciò che ne consegue impone alla scienza sociale di trovare nuove forme e nuove idee per rendere proficui i propri piani di campionamento se non, addirittura, per rendere credibile ed utile la propria attività.

1.1.3 Ricerca sociale e governo del mutamento: una breve riflessione

La *privacy*, il problema dell'accesso democratico alle informazioni, la diffusione non uniforme delle tecnologie di comunicazione e altro ancora, rappresentano dei reali ostacoli rispetto al tentativo di coniugare la teoria con la pratica di ricerca. Rimanere al passo con il mutamento dei mezzi di comunicazione è certamente una necessità²⁹, ma difficilmente questa rincorsa può diventare una soluzione alle problematiche di copertura campionaria. In altri termini, la tecnologia che si può applicare alla ricerca sociale non necessariamente riuscirà a ricostituire quell'urna che qualcuno ritiene si sia rotta. Come sempre nella storia, la scienza deve fare i conti con l'introduzione di idee, metodi e procedure che devono essere validate e controllate. Ciò permette di ridisegnare progressivamente i contorni della propria cultura e rendere in tal modo chiaro il piano informativo. Distinguere ciò che è possibile fare da quel che si ritiene corretto fare è una necessità imprescindibile della scienza, sia per la didattica sia per le sfide a cui si espone. Prima tra tutte, la sua accettazione in quanto scienza. Le problematiche sollevate dall'evoluzione tecnologica riguardano più gli aspetti culturali (metodologici) piuttosto che le trasformazioni in sé. Le vere questioni da risolvere sono riferibili alla capacità della scienza di governare i cambiamenti.

1.1.4 Tecnologia e raccolta dati

L'evoluzione tecnologica, come l'evoluzione naturale, è potenzialmente in grado di far nascere nuove forme di raccolta dati, ibridare modalità che oggi sembrano inconciliabili ed estinguerne altre. È forse questo il principale e concreto effetto della tecnologia che deve essere monitorato. Eccone alcuni esempi.

²⁹ Necessità dettata dall'esigenza di offrire adeguati strumenti di comunicazione affinché ogni unità campionaria abbia la reale opportunità e/o la più ampia libertà di esprimere il proprio parere.

1.1.4.1 Unificazione dei supporti per la raccolta dei dati

Oggi distinguiamo l'indagine, come indicato nella Figura 1.2, in base all'utilizzo di un questionario cartaceo, del pc e, recentemente, anche in base all'uso del web quale strumento per la trasmissione in tempo reale dei dati raccolti (WAPI - *Web Assisted Personal Interviewing*). Ma se il questionario cartaceo diventasse un sottotipo di un *e-book*, un documento elettronico interattivo capace di espandere pressoché a piacimento l'insieme cognitivo che sta alla base del giudizio da esprimere?

Figura 1.2: Metodi e strumenti per la raccolta dati

Supporto	Contatto		
	Telefonico	Personale	"Nessuno" (Autosomministrata)
Computer	CATI	CAPI (stand alone) WAPI (rete internet)	CAWI (rete internet)
Carta (Paper)	PATI (in disuso)	PAPI (in disuso)	POSTALE, presso uffici, ecc.
e-book multimediale	EbTI (?)	EbAPI (?)	EbWI (?)

Un *e-book* rappresenta solo in parte la trasformazione digitale di un libro classicamente inteso. Esistono esperimenti editoriali digitali che si propongono l'obiettivo di espandere il contenuto scritto del libro attraverso l'inserimento di *link* multimediali. Un oggetto, un paesaggio descritto dall'autore può essere effettivamente visto tramite delle foto digitali collegate al testo. La descrizione scenica di un evento può essere prima letta e poi "vissuta" grazie a spezzoni di un film a cui si rimanda immediatamente la visione. I personaggi citati in un romanzo storico possono essere maggiormente conosciuti grazie all'ausilio delle informazioni presenti in internet. In sintesi, grazie alla tecnologia digitale, alcuni autori stanno tentando di espandere l'insieme delle sensazioni e delle informazioni strettamente connesse alla loro opera. L'*e-book*, fondamentalmente, non solo può coinvolgere i sensi umani più adatti affinché determinati messaggi siano recepiti (udito e/o vista, almeno per ora), ma consente anche una lettura personalizzata del libro stesso. Ogni lettore è libero di espanderne i contenuti, intraprendere un percorso di

lettura piuttosto che un altro (sviluppo di trame parallele), apporre commenti, ecc., finanche interagire direttamente con l'autore.

Cosa rimane di sostanziale nella differenza tra carta, pc o pc-web se la flessibilità sulla forma espositiva del questionario, il modo di "sfogiarlo", il modo con cui si può indicare una preferenza o inserire una dato subirà queste trasformazioni? Rimane ancora importante contraddistinguere che il dato è stato raccolto con carta e penna, rispetto all'aver schiacciato un tasto del computer, fatto un click con il mouse o che la trasmissione sia avvenuta in tempo reale via internet? Se non si entra nel merito (che probabilmente sarà sempre più indistinto) delle peculiarità di ogni singolo supporto, la diversità tra le tecniche sembra veramente minuta al confronto con quello che la tecnologia consente oggi di sviluppare in termini di strumento di rilevazione. Ciò è tanto più vero oggi che esiste l'opportunità di interagire con un foglio di carta virtuale (commentare, sottolineare, ecc.) forse oltre a quanto è possibile fare con un foglio di carta reale/tradizionale. Con l'introduzione tecnologica, le prerogative dei vari supporti per l'acquisizione del dato si ibridano tanto quanto le forme della comunicazione, facendo così perdere di significato le distinzioni attuali.

1.1.4.2 Ibridazione delle tecniche di raccolta dati e ruolo del rilevatore

Il tentativo di far confluire in un'unica piattaforma software sia il CATI che il CAWI e, ancora, farlo coesistere in un'indagine con tecnica di raccolta mista, ossia con la presenza di un PAPI (a sua volta trasformato in WAPI), è una linea di sviluppo già in atto³⁰. La convergenza tecnologica rappresenterà un investimento importante nei prossimi anni e avrà come prima conseguenza l'abbattimento dei costrutti consolidati con cui oggi si identificano le diverse tecniche d'indagine. Dal punto di vista economico, essa assolverà alla necessità di superare i limiti connessi al processo

³⁰ Diverse sono le società che partendo da una soluzione CATI hanno nel tempo sviluppato una integrazione del prodotto con strumenti CAWI. Una per tutte, il prodotto BLAISE® del Central Bureau Of Statistics Netherlands che sta attualmente sviluppando una soluzione (Blaise 5) completamente orientata al web, dopo che con Blaise 4 era uscita definitivamente dal mondo (non visuale) DOS (Blaise 3).

produttivo, come ad esempio la possibilità di avere immediatamente il dato informatizzato e pronto all'elaborazione. Dal punto di vista pratico, tutto dipenderà dalla condivisione degli strumenti di comunicazione con il soggetto osservato.

E la figura del rilevatore che funzione ha e avrà? Se non ha il ruolo di semplice osservatore³¹, il rilevatore è colui che potenzialmente (ma non necessariamente) offre un supporto tecnico alla compilazione³² ed, eventualmente, qualifica all'intervistato il punto di vista con il quale il ricercatore vorrebbe fosse valutato un determinato tema d'indagine. Non è per nulla scontato, infatti, che la semplice lettura di una domanda (l'iniziale stimolo percepito dall'intervistato, stimolo che potrebbe anche semplicemente essere una immagine grafica o la registrazione vocale della domanda) sia sufficiente affinché tutti gli intervistati interpretino allo stesso modo una situazione da giudicare, un fatto da quantificare in frequenza, ecc. La domanda produce uno stimolo che è esso stesso fonte di variabilità estranea rispetto al fenomeno misurato. Per questo motivo, il quesito va formulato attentamente, promosso in modo identico a tutti gli intervistati e, semmai ci fossero problemi interpretativi, rielaborato (dal rilevatore o con altre informazioni presenti sul questionario) con la stessa prerogativa di unicità e condivisione della semantica desiderata dal ricercatore³³. Al rilevatore potrebbe essere dato il compito di trasferire all'intervistato un determinato e più limitato punto di vista con il quale si vuole (ad esempio) analizzare un argomento d'interesse che notoriamente presenta diverse sfaccettature, ossia un tema complesso e

³¹ Ad esempio un'indagine che prevede il conteggio di quanti salgono e scendono a una fermata del bus.

³² Come ad esempio il semplice richiamo rivolto all'intervistato affinché esprima il proprio giudizio in coerenza con le risposte previste dalla domanda.

³³ È utile ricordare che uno dei punti di forza a cui si appellano coloro che sostengono la bontà del metodo di perseguire l'analisi dei fenomeni sociali attraverso la collezione di informazioni proveniente da internet (*sentiment analysis/opinion mining* attraverso i *big data*), sta proprio nell'assenza del questionario. A tal proposito si veda l'accento fatto a questo aspetto nel già citato articolo di www.youtrend.it. Ma attenzione! Ciò non significa che non si rendano necessari cospicui lavori di standardizzazione della semantica e di analisi della pertinenza al tema d'indagine di quanto viene raccolto in termini di espressioni verbali (dati) in internet.

multidimensionale. Nello specifico, domanda per domanda, il rilevatore è tenuto a fornire il punto di vista imposto dal ricercatore e, proprio per questo, egli riceve un'adeguata formazione.

Il modo con cui è formulato un quesito, le ulteriori informazioni ritenute utili a specificare la tematica e l'intervento di un rilevatore a mediazione del processo cognitivo che l'intervistato deve svolgere, rappresentano una fonte di variabilità estranea al fenomeno indagato. Obiettivo del ricercatore è quello di minimizzare se non annullare tale variabilità e i tipici effetti di *question-wording* (Bernardi (2005), Bernardi e Tuzzi 2005). È anche per questo motivo che a volte si preferisce non prevedere affatto il rilevatore e far compilare il questionario all'intervistato in solitaria (autosomministrazione)³⁴. In sostanza, il distinguo tra le tecniche d'indagine prevede una situazione dicotoma circa il rilevatore: presente – non presente. L'espansione della telefonia fissa ha di fatto creato una rielaborazione di questo stato. La presenza del rilevatore è sempre prevista ma è mediata dal telefono (il CATI). Il divario tra l'esserci (direttamente o mediaticamente) o non esserci, tuttavia, è rimasto ampio. Ma rimarrà sempre così? Sia per le componenti tecniche dell'intervista sia per gli aspetti di contenuto? Dal punto di vista dei contenuti, si è precedentemente parlato dell'*e-book* e delle sue potenzialità multimediali per dare l'idea di come si possano ibridare le peculiarità elaborative del pc e quelle documentali tradizionali del libro. Si è messo in luce che l'obiettivo è ampliare e articolare la cognizione di ciò che si sta leggendo (immagini, video, approfondimenti tematici, ecc.). Questi strumenti multimediali potrebbero essere considerati i mezzi a cui il rilevatore (o l'intervistato) può attingere al fine di qualificare al meglio il tema sottoposto a valutazione (la domanda). Passando al lato tecnico, potremmo dire che l'espansione di tali possibilità non incide sulla discrezionalità circa il loro

³⁴ Molto più spesso il rilevatore non è previsto perché rappresenta un costo di produzione. Al ricercatore spetta quindi il compito di valutare il rapporto tra costi (economici e metodologici) e benefici indotti da tale figura.

utilizzo, così come rimane da approfondire il previsto o assente ruolo del rilevatore. Anche qui la tecnologia ci può venire incontro. Se allarghiamo per un attimo la visuale relativamente ai supporti multimediali con cui oggi si entra in contatto, la presenza o meno di un rilevatore non è necessariamente una condizione da precisarsi necessariamente a priori. Egli, grazie alla tecnologia, potrebbe “materializzarsi” o “scompare” a seconda della volontà dell’intervistato o in ragione della presenza, assenza e/o intensità di determinati fattori “ambientali”. La barriera dello spazio dal punto di vista tecnologico è già stata abbattuta da molto tempo. La possibilità di attivare *l’human coaching* tra computer e computer è, ad esempio, una realtà consolidata³⁵. A questo proposito, dobbiamo anche ricordare la tecnologia che permette di “leggere” all’intervistato (con programmi di *text to speech*) quanto a volte viene annotato sul questionario in termini di informazioni e approfondimenti sulla domanda. Programmi, oggi combinati con la messa a punto di *avatar virtuali*³⁶, capaci di aiutare l’intervistato nel percorso di compilazione del questionario. La combinazione tra i tempi di risposta, la richiesta esplicita di ulteriori delucidazioni, piuttosto che altri eventi informatici, possono generare dei veri e propri stati interattivi a parziale o completa sostituzione del rilevatore.

Tutto ciò, pur configurandosi come una scienza fantasiosa fin da ora attuabile, elaborata con strumenti già presenti e semplicemente assemblati allo scopo, dà l’idea di come le peculiarità dei diversi metodi d’indagine possano essere oggi sostanzialmente riviste. L’interazione tra intervistato e rilevatore (reale o virtuale che sia) non può più essere considerata una variabile a stati fortemente discreti, semmai – direi – quasi continua. Anche solo questo aspetto, tutto ciò che questo implica dal punto di vista degli

³⁵ Nell’assistenza tecnica offerta ai propri clienti da molti produttori di software, è d’uso predisporre delle funzionalità dette di *human coaching*, per condividere lo schermo del pc del cliente durante una videochiamata tra le parti.

³⁶ Immagini per rappresentare se stessi nella realtà virtuale.

effetti di *question wording*, rende precaria la suddivisione delle tecniche d'indagine a cui oggi facciamo riferimento.

Chi è tentato di pensare che quanto detto non sia già materia di studio e di sperimentazione o che si configuri come uno scenario non praticabile, può esaminare la tecnica CARI (*Computer Audio – Recorded Interviewing*)³⁷ utilizzata dal *Survey Research Center – Institute for Social Research* dell'Università del Michigan; oppure i lavori svolti e gli articoli prodotti dal *U.S. Census Bureau, Statistics Canada, RTI International*³⁸. Molti sono gli esempi di tecniche già sperimentate, tra cui valutazioni d'impatto sulle stime ottenute attraverso l'ausilio o meno di *avatar* in fase di rilevazione dei dati e altri supporti interattivi.

La sensibilità dei temi trattati e la necessità di trovare (direi) un'armoniosa globalità di condizioni al fine di assicurarsi la qualità del dato (Dillman et al, 1995, Dillman et al, 2009), rappresentano situazioni con cui confrontarsi a livello di metodo. La tecnologia esiste e sempre più darà spunto a nuove reinterpretazione degli strumenti e delle modalità di somministrazione. Dal punto di vista metodologico bisognerà capire se l'evoluzione tecnologica porterà a sommare i pregi e ad abbattere alcuni difetti delle attuali tecniche d'indagine, oppure il contrario. È corretto tuttavia far notare che alla base di questa prospettiva di cambiamento non c'è solo la disponibilità attuale o prossima ventura di una tecnologia che permetta il superamento di ciò che oggi identifichiamo come fattori distintivi (a volte limiti/difetti, a volte pregi) di una tecnica/modalità

³⁷ Quando si parla di audio-registrazione delle interviste, si deve intendere la registrazione vocale sia delle risposte fornite dall'intervistato, sia della parte relativa alla domanda. Le indagini avvengono attraverso l'utilizzo di "agenti sintetici" (detti I.V.R. – *Interactive Voice Response*) capaci di "leggere" qualsiasi testo scritto, anche modificato interattivamente nel corso dell'indagine stessa. L'IVR si caratterizza come "agente" in quanto è in grado non solo di fungere da registratore ed espositore vocale di un testo, ma anche di interpretare a livello vocale quanto gli viene detto verbalmente o quanto viene digitato su una tastiera (interpretazione del segnale digitale). Basilari esempi in tal senso sono le richieste di far partire una chiamata dal nostro telefonino semplicemente pronunciando il nome della persona presente in rubrica (http://en.wikipedia.org/wiki/Interactive_voice_response).

³⁸ Per la presentazione della tecnica CARI da parte di RTI international si veda il link <http://www.rti.org/brochures/cari.pdf>. Circa gli effetti sul tasso di risposta di diverse tecniche d'indagine, tra cui l'IVR, si può fare riferimento all'articolo di Dillan (Dillan et al. , 2009)

d'indagine, quanto la sua accettazione "strumentale" e la conseguente diffusione presso la popolazione. Il *digital divide*, infatti, è notoriamente connesso a diversi fattori economici e culturali che potrebbero non trovare soluzione o attenuarsi nel breve-medio periodo. La stessa velocità con cui avvengono le trasformazioni tecnologiche costituisce un impedimento all'omogeneizzazione e/o alla convergenza degli strumenti di comunicazione e di raccolta dei dati. Facendo capo all'idea che la tecnologia, capace di cose inimmaginabili fino a poco tempo fa, non ha il potere di contrarre le cause del *digital divide*, si può pure essere favorevoli all'ipotesi che permanga uno scenario che vede la presenza di diverse modalità e mezzi di comunicazione. Di conseguenza, anche di strumenti di rilevazione.

1.1.4.3 Lo strumento di rilevazione: da questionario a gestore di processo

Fino a questo momento si è focalizzata l'attenzione sull'impatto della tecnologia nel distinguere e nello sviluppo delle diverse tecniche d'indagine. In particolare, si è puntata l'attenzione sul fatto che, lentamente ma in modo inesorabile, il personal computer è riuscito ad erodere ed inglobare le peculiarità dei vari strumenti di raccolta dati (il questionario cartaceo), così come ha potenzialmente reso sempre più esile lo spazio che separa una tecnica d'indagine dall'altra. Anche in questo senso è plausibile parlare di convergenza tecnologica (cfr. par. 1.1.1.2). Le righe e le colonne della Figura 1.2 dovrebbero essere riunite, così da configurare una sola cella. L'avanzamento tecnologico infrange infatti lo stereotipo dei diversi metodi di raccolta dati e la varietà delle possibili soluzioni è tale da imporre una qualificazione dettagliata del procedimento d'indagine utilizzato.

Trattando questi argomenti, inevitabilmente si è parlato anche di strumenti d'indagine, quasi sempre alludendo al questionario, attraverso il quale si definiscono e si annotano le misure (modalità) relative all'argomento specifico analizzato (domande). È auspicabile che le domande e le risposte siano formalizzate in modo adeguato, o per meglio dire, personalizzate in ragione di risposte fornite in precedenza o in base a

situazioni conosciute a priori (ad esempio il genere della persona contattata). Altre volte è utile proporre alcuni quesiti e/o le relative modalità di risposta in forma casuale, così da favorire la massima dispersione dei “rumori di fondo” che sono frutto di azioni negligenti da parte dell'intervistato o del rilevatore (l'imputazione di un codice qualsiasi pur di proseguire l'intervista). È altrettanto frequente che un intervistato debba rispondere ad alcune domande piuttosto che ad altre e che tale percorso di compilazione sia la risultante di molteplici stati che si rendono palesi in itinere (attività di *screening e routing* del questionario). La conformazione dello strumento d'indagine, quindi, rappresenta per il ricercatore una necessità, strettamente legata alle misure rilevate e alla qualità delle stesse, cioè alla chiarezza d'intenti proposta all'intervistato e alle distorsioni che ne derivano. Recenti studi (Cocco e Tuzzi 2012) evidenziano, tra l'altro, differenti effetti di *question wording* in ragione dello strumento tecnologico di rilevazione.

Ciò che differenzia il questionario informatizzato da quello cartaceo è l'opportunità di adeguare lo strumento in itinere sulla base di eventi informatici/informativi che non necessariamente appartengono all'insieme delle risposte acquisite. Ad esempio, i tempi di risposta e la richiesta di delucidazioni possono attivare ulteriori eventi³⁹, anche del tutto estranei al fine stesso dello strumento di rilevazione. Ciò che ci preme sottolineare è che, inglobando la capacità elaborativa del pc, lo strumento di rilevazione (il questionario informatizzato) diventa una “macchina” che permea e raccorda quasi tutti gli atti inerenti la raccolta dati: dal controllo circa la validità dei valori inseriti, alla verifica delle linee di campionamento, ecc. Tale arricchimento funzionale rende pressoché impropria la qualificazione del questionario quale “semplice” strumento di misurazione. D'altra parte, è intuibile che l'interfaccia grafica con la quale si acquisisce l'informazione non costituisce il differenziale tra un'applicazione CATI/CAWI e un

³⁹ In precedenza, ad esempio, si è ipotizzata la possibilità di attivare una linea di comunicazione con un tecnico/rilevatore a supporto della compilazione del questionario.

semplice programma gestionale. Un software per la raccolta dati si qualifica, più di tutto, in ragione della capacità di adattarsi agli “ambienti” e strumenti con cui opera (input dati, *smartphone*, indagini personali, telefoniche, web, ecc.) e della pervasività, ovvero rispetto alla flessibilità nel raccordare le diverse attività, dalla modalità di trattamento delle unità di rilevazione⁴⁰ fino al supporto e/o autonomia nella gestione dei tipici processi decisionali. Non a caso, la programmazione di un questionario elettronico (CATI e CAWI) è da tempo arricchita dal concetto informatico di azione su evento.

Valutiamo alcune concrete innovazioni che delucidano al meglio quanto affermato.

Si pensi, come primo esempio, all'informatizzazione delle tecniche d'indagine N.G.T. (*Nominal Group Technique*) e/o Delphi (VandeVen e Delbecq, 1974). Il tipico obiettivo di configurare uno scenario unitario rispetto ad un argomento/evento o di rendere palese la presenza di opinioni divergenti attraverso il confronto di numerosi punti di vista e competenze, non viene spogliato delle fasi relative al contraddittorio tra i “testimoni”, ma reso veloce grazie alla presentazione *just in time* dei giudizi numerici (statistiche inerenti le posizioni di valore dei partecipanti) e dei contenuti verbali espressi (spiegazioni delle proprie posizioni di valore).

Un ulteriore esempio è quanto si è riusciti a raggiungere nel campo della standardizzazione dei processi di raccolta dati. Se partiamo dal presupposto che una somministrazione troppo lenta o troppo veloce di un questionario è spesso fonte di possibili errori extracampionari, allora è possibile valutare la distribuzione dei tempi medi di intervista di ogni rilevatore rispetto a quelli della ricerca nel suo complesso così da isolare i rilevatori e/o le interviste che si collocano al di fuori di un *range* ritenuto “fisiologico”. Il comportamento del rilevatore può essere

⁴⁰ Al netto degli appuntamenti, la programmazione dei contatti deve favorire l'equiprobabilità di rispondere a tutte le unità presenti nel campione. Per far questo, ad esempio, nei software CATI è prevista la possibilità di alternare giorni ed orari di chiamata.

conseguentemente valutato e modificato praticamente in tempo reale. Ciò aiuta la standardizzazione della rilevazione al fine di contrarre la varianza estranea al fenomeno indagato (Campostrini, 1990). Questa prassi può essere impiegata non solo per i tempi di somministrazione, ma anche per altri aspetti fondamentali dell'indagine. Tassi di rifiuto troppo elevati, ad esempio, potrebbero accompagnarsi a bassi tempi rispetto ai quali il rilevatore mediamente chiude i contatti giudicandoli come dei rifiuti a partecipare. Ciò permette al supervisore di ribadire all'intervistatore l'importanza di promuovere al meglio la ricerca; aiutandolo a variare il modo di presentarsi e spronandolo a insistere maggiormente e con più argomentazioni durante la conversazione con il soggetto da intervistare.

In riferimento al piano di campionamento, invece, è possibile definire in tempo reale la numerosità campionaria ottimale per ogni strato in ragione di una raggiunta stabilità della varianza del fenomeno che si sta indagando, così come proposto da Neyman per la ripartizione ottimale in un campione stratificato (Diana e Salvan, 1993).

Un altro interessante aspetto è quello rappresentato dalle risposte a tempo. Esistono argomenti di indagine in cui è importante valutare il differenziale tra il pensiero consapevole-razionale (derivante da un percorso cognitivo che molto spesso viene poi proposto all'intervistato) e quello emotivo-d'impulso o di preconcetto. L'ausilio del computer facilita di molto questi esperimenti, permettendo di attivare dei veri e propri cronometri che scandiscono il tempo utile alla risposta. In alcuni software CAWI, Limesurvey^{®41} ad esempio, è possibile stabilire addirittura il tempo minimo di risposta, di grande utilità quando le domande sono sottoposte al rispondente "in batteria" e quindi confondibili l'una con l'altra.

Si può proseguire negli esempi ricordando la funzione di *drag and drop* in grado di prelevare oggetti grafici (frasi, parole o disegni) e spostarli

⁴¹ <https://www.limesurvey.org/>

fisicamente da un punto all'altro dello schermo, così da metterli in ordine nello stesso modo in cui si opererebbe manualmente nello spazio reale.

L'introduzione della tecnologia, tra l'altro, non riguarda solo la multimedialità, la *gamefication*⁴² introdotta durante la raccolta dei dati e le caratteristiche tipiche dell'elaborazione, ma è pervasiva anche in altre discipline come la geografia. Gli attuali sistemi di posizionamento al suolo (GIS – *Geographic Information System*) permettono di verificare, con un semplice *smartphone*, i percorsi effettuati dal rilevatore che deve presidiare un'area geografica o recarsi in uno specifico luogo a svolgere un'intervista. E se si volesse controllare il grado di "sensibilità" dell'intervistato rispetto ai temi trattati? O, ancora, si volessero ricevere informazioni circa la qualità dell'interazione tra intervistato e intervistatore? In questi casi alcuni software CTI (*Computer Telephony Integration*) sono già stati dotati di moduli per il controllo *real time* dello stress vocale presente durante una conversazione telefonica⁴³.

Appare evidente che la tecnologia trasformerà tanto il controllo quanto le modalità di somministrazione dello strumento di rilevazione. Cambieranno le opportunità concrete di svolgere determinate indagini, così come si modificherà il costrutto logico con il quale distinguiamo oggi le diverse tecniche d'indagine⁴⁴. I percorsi scelti potranno rientrare o meno nei canoni della cultura della ricerca e, per questo, ancora una volta è importante affermare che le prassi e gli strumenti adottati, frutto di

⁴² *Gamefication* e *Surveytainment* sono due termini che Couper (2013) usa per descrivere la tendenza a trasformare il momento della raccolta dei dati (il questionario elettronico) in un momento ludico, di gioco, d'intrattenimento. Lo scopo è quello di tenere più a lungo possibile il rispondente al pc (o allo *smartphone*) così da incrementare i tassi di adesione. Il rischio collegato a questa tendenza, dice Couper, è quello di rendere insignificante (*trivialize*) l'attività che si sta conducendo. Egli afferma che "noi [ricercatori] vogliamo persone che prendano quello che facciamo (e che cosa chiediamo loro di fare) seriamente". Altrimenti " *is like putting lipstick on a pig*".

⁴³ Si possono trovare informazioni in http://en.wikipedia.org/wiki/Voice_stress_analysis/, mentre un prodotto commerciale è il software X13-VSA della compagnia omonima (X13-VSA Ltd - www.Lie-Detection.com).

⁴⁴ Con l'avvento nello scenario di ricerca dei *big data*, si arriva addirittura ad ipotizzare che i dati relativi ai comportamenti delle persone e le loro idee saranno presenti nel web e noti a tutti, così da annullare il presupposto stesso dell'azione di indagine e, di conseguenza, pure la necessità di predisporre gli strumenti di rilevazione.

molteplici combinazioni, devono rimanere sempre e comunque al centro dell'attenzione del ricercatore, che potremmo dire, non può e non deve diventare agnostico rispetto alle metodologie e alle fonti dei dati che analizza. Esiste oramai una letteratura assai ampia, che non può essere ignorata, a riprova dei cospicui effetti sui tassi di risposta, sulle distorsioni nelle stime, ecc. che si possono ingenerare utilizzando strumenti di rilevazione diversi, questionari diversi, domande differenti ecc. ; ambiti di studio riunibili sotto la voce di effetti di *question wording*⁴⁵.

Tale rischio, al contrario, è concreto e, a mio avviso, dovuto alla storica evoluzione in comparti specializzati che coinvolge il settore della ricerca così come tutti gli altri sistemi produttivi. Il processo di concentrazione strumentale, come già ribadito, prende vigore solo in parte da esigenze strettamente metodologiche; molto più facilmente da esigenze di efficienza produttiva connessa – semmai – a esigenze di carattere metodologico. Qualcuno potrebbe considerare questi ambiti degli elementi riguardanti l'economia e, di conseguenza, per certi versi estranei alla metodologia della ricerca. Il punto di vista con il quale si è affrontato fino ad ora questo argomento, al contrario, considera le trasformazioni produttive e strumentali un elemento conseguente a ciò che è già avvenuto e nello stesso tempo condizionante le trasformazioni ulteriori che si producono tanto nella ricerca applicata, quanto nelle idee scientifiche a cui questa fa riferimento. Se consideriamo che la tecnologia scelta, l'organizzazione e la prassi sono elementi che incidono sul prodotto intermedio (la qualità del dato), allora si può paragonare la ricerca applicata a qualsiasi altro processo produttivo e, di conseguenza, processo in evoluzione. Uno sguardo economico a come i sistemi produttivi si evolvono e come trasformano l'ambiente socio-culturale è sembrato quindi un tema

⁴⁵ Citiamo a riguardo gli studi di Banks (Banks et al, 2003), Bishop (Bishop et al. 1988), Couper (2000, 2008), Weber (2008) e Kreuter (Kreuter et al. 2008)

appropriato nella trattazione dell'impatto che la tecnologia ha nella ricerca sociale. A questo tema è dedicato il prossimo capitolo.

1.2 La produzione di informazione e le trasformazioni del comparto della ricerca

1.2.1 Premessa

Quando si implementa un'indagine si inseriscono elementi generalmente estranei al piano teorico di riferimento. Bisogna tener conto dei tempi con i quali si riuscirà a organizzare la struttura di rilevazione, quelli necessari alla predisposizione degli strumenti d'indagine, quelli legati alla validazione di detti strumenti e, non ultimo, considerare la compatibilità tra i tempi di raccolta dati e l'obsolescenza che questi accumulano nella successiva fase d'analisi. I costi, d'altra parte, così come la possibilità di rendere simultanee le varie operazioni, sono ulteriori elementi da governare. Dare attuazione ad una ricerca significa programmare una sequenza di parcellizzate attività interconnesse al pari di quel che succede in ogni altro comparto produttivo. Se la creazione dell'informazione, che trae spunto dai dati, può essere considerata a tutti gli effetti un processo produttivo, allora la qualità del prodotto finale non può che dipendere dalla messa a punto del processo stesso, così come l'assetto del comparto produttivo non può essere considerato un fattore estraneo. Se così non fosse, significherebbe che qualsiasi produzione sarebbe possibile ed equamente efficiente in qualsiasi organizzazione, territorio e condizione infrastrutturale di contesto. Analizzare la trasformazione dei processi produttivi, comprenderne le origini e gli effetti diventa perciò un passaggio obbligato per acquisire consapevolezza circa l'evoluzione del comparto della ricerca sociale e valutare la qualità dell'informazione prodotta.

1.2.2 La natura dei costi e l'evoluzione dei processi produttivi

Quando si realizza un'indagine, parte dei costi sono dettati dal numero di interviste preventivate; altri, invece, esulano da ciò e rappresentano un fattore fisso collegato prettamente all'impianto metodologico. Entrambi questi costi sono detti diretti. Ogni produzione, anche parziale e quindi inserita in un sistema produttivo più complesso, comprende quantomeno i

costi diretti. Non bisogna poi dimenticare i costi indiretti, tra i quali quelli denominati transazionali, ovvero le spese connesse alla necessità di organizzare e mettere in relazione le parcellizzate attività produttive.

Un processo di raccolta dei dati può essere visto come un'attività industriale: una sequenza ordinata e organizzata di frammenti (attività e/o produzioni parziali) che, una volta assemblati, definiscono il prodotto finale. Questo è ciò che differenzia una produzione individuale/artigianale da una industriale. Se tali spezzoni del processo non sono correttamente assemblati, il rischio è quello di costruire un oggetto malfunzionante. Per la ricerca sociale, dove il prodotto è pressoché immateriale, è più difficile controllarne "il corretto funzionamento". Ciononostante, non è saggio dimenticarsi che anche nella ricerca si è di fronte ad un processo industrializzato o che si sta industrializzando o che si vuole industrializzare.

La specializzazione produttiva è orientata (funzionale) all'abbattimento dei costi diretti, mentre l'organizzazione (e il relativo costo) rappresenta sempre più l'elemento che qualifica il processo nel suo complesso. Tanto più il processo si parcellizza/specializza, tanto più diventa importante l'organizzazione produttiva e diventano probabili gli incrementi dei costi transazionali. Gli oneri organizzativi (sia strumentali che di coordinamento vero e proprio), fanno pienamente parte dei costi dell'informazione (il prodotto). La qualità del prodotto, di conseguenza, include e non può fare a meno dei costi transazionali. Si capisce quindi per quale motivo si tenda a far diventare gli strumenti d'indagine i veri gestori del processo di produzione dei dati. Tramite la tecnologia, infatti, si cerca di abbattere sia i costi diretti sia quelli organizzativi.

Teoria, metodo e costi, nella ricerca applicata si legano in modo indissolubile e alimentano reciproche pressioni atte a modificare i corrispondenti spazi d'influenza. A partire dalla natura dei costi si possono declinare due ambiti d'analisi tra loro interconnessi: quello della

trasformazione delle società che si occupano di ricerca applicata e quello relativo alla qualità dei dati. Entrambi fanno riferimento esplicito all'organizzazione delle strutture produttive, vale a dire alla crescente specializzazione delle produzioni collegate alla creazione dell'informazione (il prodotto di ricerca). Se la prima delle due, quella riguardante gli attori del comparto produttivo, potrebbe sembrare indifferente rispetto al dibattito metodologico, la seconda di certo lo investe in pieno.

1.2.3 Qualità nei processi di raccolta dei dati

Per affrontare in modo pragmatico gli aspetti dell'organizzazione dei processi produttivi nell'ambito della ricerca, ci si dovrebbe chiedere a cosa sono dovute le evidenti differenze nelle proposte economiche relative alla stessa richiesta/prodotto. I margini di guadagno possono esserne la causa. I costi del personale (spesso comuni a tutti i produttori all'interno di una nazione) un'altra. Lo possono essere le tecnologie o le economie di scala dettate dai volumi⁴⁶. Nella ricerca sociale, data l'immaterialità di quel che si acquista e la difficoltà a controllarne il "buon funzionamento", ci si dovrebbe chiedere, però, se tra le pieghe del processo produttivo si materializzano risparmi tali da mettere a repentaglio la qualità stessa dell'indagine.

I ricercatori spesso considerano l'impianto organizzativo e di produzione "solo" (pur giustamente) come fonte di possibili errori

⁴⁶ Diverse sono le fonti delle economie di scala, considerate da sempre il motore del processo cumulativo della ricchezza e del progresso economico. L'eccessiva parcellizzazione del lavoro, tuttavia, può assumere un carattere di diseconomia quando i costi transazionali superano la componente positiva derivante dall'aver industrializzato (parcellizzato e specializzato) il processo di produzione. È quanto emerge da alcune analisi organizzative basate su altrettante teorie riguardanti i costi transazionali (Williamson, 1987). Si tratta di studi focalizzati alla comprensione della perdita di competitività e/o di crisi economica che molte realtà del tessuto produttivo in generale stanno vivendo soprattutto in questo periodo storico. La risposta economica si è concretizzata nell'internazionalizzazione e nella flessibilità, cioè nella ricerca di componenti di competitività attraverso la diminuzione dei costi di produzione mediante il trasferimento della produzione in paesi con un minore costo del lavoro. Ciò ha creato spesso un allontanamento dei luoghi di produzione dai luoghi di consumo, elemento scarsamente tenuto in considerazione nel pensiero economico classico (liberale) o meglio, ritenuto connaturato all'idea stessa di libero mercato. Se la politica è tenuta ad interrogarsi su questi aspetti della società e della sua evoluzione, la metodologia della ricerca è tenuta a comprenderne l'impatto sui propri meccanismi di produzione, perché su questi aspetti sta spesso la differenza qualitativa del prodotto.

extracampionari. Per la tecnica CATI, ad esempio, si può definire una lista alquanto corposa di situazioni/elementi fonte d'errore extracampionario, tra le quali:

- l'utilizzo degli *spare-panel* telefonici⁴⁷;
- la mancanza di strumenti per valutare la standardizzazione del comportamento dei rilevatori (tempi di somministrazione, tassi di rifiuto, ecc.);
- l'insufficiente supervisione (scarso supporto offerto ai rilevatori nel dirimere i dubbi collegati alla natura stessa delle informazioni raccolte);
- la durata complessiva dell'azione d'indagine (Keeter, 2000)
- i controlli (*real time* e non) delle digitazioni dei rilevatori incoerenti o poco probabili;
- la complessità visiva e di gestione del questionario informatizzato (facilità d'uso dello strumento CATI e impostazione video del questionario);
- la comprensione telefonica delle domande (lunghezza, linguaggio utilizzato, ecc.);
- il sistema di gestione della lista dei contatti (equiprobabilità offerta ad ogni numero di telefono, cioè modalità e tempi di trattamento dei contatti).

Analogamente, nelle indagini *face to face* è noto che il comportamento del rilevatore gioca un ruolo ancora più forte sulla qualità del dato. L'essere isolato rispetto alla struttura di ricerca, infatti, favorisce comportamenti che determinano errori sistematici.

Se la stessa solerzia fosse utilizzata tanto nel determinare le cause degli errori quanto nella definizione del processo produttivo, probabilmente i differenziali di prezzo avrebbero maggiori *chances* d'essere compresi. In altre parole, qualora ci fosse maggiore attenzione nel

⁴⁷ L'adozione di queste pratiche è ben nota e assai diffusa (Di Gioia, 2009), tuttavia è quasi impossibile trovare un bando di gara pubblico che faccia esplicito divieto di utilizzo di queste pratiche così fortemente connesse agli errori extracampionari.

declinare nella sostanza i processi desiderati (ad esempio nella richiesta di un preventivo), si limiterebbero i danni provocati alla ricerca dall'opportunità fornita ai vari istituti di "saltare a piè pari" determinati costi e le relative attività utili alla salvaguardia della qualità. Nella ricerca sociale, l'informazione che si trae dal dato non è indifferente dal come viene acquisito⁴⁸. Se nulla di preciso è richiesto e tutto si concentra nel perseguire un risultato numerico campionario, quale comparazione di costo è possibile e quale conoscenza degli errori extracampionari si può avere? Presumere che le prassi di lavoro, le tecnologie adottate e le organizzazioni siano uguali per qualsiasi produttore è un errore. Significa pensare che esista una produzione oramai standardizzata o la possibilità di controllare la qualità della fornitura ex-post ed eventualmente contestarla. Il controllo sulla qualità del dato raccolto, invece, è un'ulteriore costo che nessuno quasi più si accolla se non altro per il fatto che ciò altera in modo consistente la programmazione temporale delle attività.

Si potrebbe obiettare che le tecnologie, le prassi e l'organizzazione del processo di raccolta dei dati potrebbero anche risultare, in alcune o svariate situazioni, indifferenti rispetto all'adeguatezza del risultato ottenuto. Tuttavia, così come lo è l'adozione della casualità nella scelta dei rispondenti, anche questi elementi dovrebbero essere attentamente valutati se non, addirittura, considerati un prerequisito. Quantomeno si applicherebbe un principio di cautela! Prendiamo quale esempio il problema della copertura campionaria e, in particolare, la possibilità che un campione si basi su una limitata parte dell'universo di riferimento⁴⁹. Come è noto, rispetto ad una caratteristica da stimare, la quantità di distorsione a cui si va incontro in questi casi è strettamente collegata al differenziale esistente tra quanti sono effettivamente contattabili e quanti non lo sono. In questo senso le stime prodotte con la tecnica CATI sembrano reggere

⁴⁸ Si potrebbe dire che la ricerca sociale è paragonabile più alla preparazione di una gustosa pietanza piuttosto che al processo produttivo di una fonderia. Gli ingredienti tanto quanto il metodo di preparazione qualifica ciò che è servito a tavola.

⁴⁹ In questa condizione si trovano tutti i campioni basati sulle pagine bianche (elenchi telefonici).

ancora alla prova dei fatti. Ma per quanto ancora? Eppure la casualità è l'unica prassi che ci consente di controllare (misurare) l'errore di stima. In caso contrario, andrebbe bene qualsiasi prassi di campionamento, anche quella non probabilistica. Similmente, un'adeguata copertura della popolazione ci rassicura circa la possibilità di rintracciare tutte le fonti di variabilità di un fenomeno, altrimenti ci si potrebbe limitare a contattare un gruppo di amici e conoscenti.

Rafforzare la cultura della ricerca, nel senso di definire le "pratiche migliori", rappresenta un percorso parallelo a quello teorico da cui decisamente non si può sfuggire. Il dibattito esperienziale è fondamentale per dotarsi di strumenti efficaci e capaci di mettere in relazione il dato raccolto alla qualità della prassi seguita e quindi al costo collegato. Tutto questo diventa ancora più importante tanto più la tecnologia rende sempre meno standardizzate le tecniche di rilevazione. Descrivere precisamente il processo produttivo è forse l'unico modo per evitare il tentativo di mascherare i costi transazionali al fine di ottenere dei risparmi non esattamente frutto di una specializzazione, di un'innovazione o di economie di scala.

Ci si potrebbe chiedere se lo sforzo di perseguire le "pratiche migliori" sia attuato nella ricerca sociale applicata, se i percorsi di certificazione (ISO) rappresentino veramente un passo in avanti e quanto il mondo scientifico e accademico partecipi e abbia partecipato a definire canoni e standard della certificazione ISO. Ciò che è certo è che l'analisi dei processi produttivi rappresenta un costo per gli acquirenti. Costo che con la certificazione si cerca di scaricare sul produttore o che, con dei surrogati, alcuni produttori tendono a contrarre a volte in modo incomprensibile e spesso del tutto inefficace. Facciamo un esempio.

Per quanto riguarda la raccolta dei dati tramite il CATI, l'Istat e altri organismi di ricerca scientifica hanno mutato nel tempo gli elementi precludenti la partecipazione ai propri bandi di gara. Quando

l'investimento in tecnologia ha rappresentato un primo tentativo (metodologicamente corretto) di contenere i costi, erano spesso (e a volte solo) richieste infrastrutture tecnologiche particolari, come un centralino con determinate funzionalità/potenzialità, un numero minimale di stampanti (laser per la precisione), una linea fax e tutto un'insieme di attrezzature che sembravano copiate dal conto patrimoniale di un'azienda piuttosto che strettamente legate alla richiesta del prodotto acquistato. Quanto era chiesto in termini tecnici, infatti, spesso non aveva proprio nulla a che fare con il processo di produzione o, quantomeno, non erano mai chiarite le ridondanze strumentali e/o le loro funzioni all'interno del ciclo produttivo. Qualche maligno poteva addirittura pensare che il tutto servisse per riuscire a pilotare l'appalto stesso! In realtà (o forse), era il semplice tentativo di cautelarsi delegando la rilevazione a quelle società che potevano, in ragione degli investimenti tecnologici fatti, offrire un buon prezzo a fronte di una buona prassi di lavoro. Successivamente, quando i *call center* (con i loro enormi apparati tecnici sviluppati per il telemarketing e per la gestione dei servizi clienti) hanno "invaso" il mercato (a loro culturalmente estraneo e per il quale non hanno spesso né competenze né strumenti specifici adeguati), i grandi istituti pubblici di ricerca hanno mutato atteggiamento. Agli stessi (più o meno) prerequisiti presenti in passato, hanno assommato l'obbligo di avere presenti nella struttura di rilevazione un numero consistente di figure professionali (dal responsabile scientifico/di progetto fino ad un congruo numero di intervistatori preventivamente individuati) che avessero una certificata esperienza sul campo e particolari titoli di studio. In alcuni bandi di gara si è arrivati a definire fino a 7-9 figure professionali a livello di ricercatori senior e qualche informatico, senza che tutte queste figure avessero un qualche ruolo di sostanza nell'attività richiesta riferibile, ad esempio e semplicemente, all'esecuzione di 10-20 mila interviste CATI. Attività per cui una media struttura di rilevazione impiega al massimo un mese di lavoro e

utilizza uno o forse due supervisori e un ricercatore alla bisogna. Alcune perplessità sorgono spontanee:

- che senso ha richiedere una precisa presenza strumentale (un fax, due e non una stampante laser, ecc.) se questa non ha nulla a che fare con lo stretto ciclo produttivo?
- che senso ha porre come vincolo la presenza di centinaia di linee telefoniche e/o figure professionali se il lavoro può essere svolto anche da una struttura più minuta dal momento che non c'è bisogno di costruire il questionario (viene fornito dal committente), i dati non sono elaborati in loco e il tutto si esaurisce nel raccoglierli?

ma soprattutto....

- come si può trascurare il fatto che gli istituti di ricerca lavorano su spare-panel telefonici? Che utilizzano strumenti di raccolta dei dati e di controllo sull'operato dei rilevatori, orari e organizzazioni e altro ancora del tutto diversi e che ciò incide sulla qualità del dato?

Il tentativo di surrogare l'analisi del processo con elementi infrastrutturali e di contorno non è certo la pratica migliore per tutelare la qualità! C'è invece sempre più la necessità di entrare nel merito dei processi stessi. Magari definendoli prima e con particolare precisione. La corretta e dettagliata descrizione del progetto di ricerca e del processo di raccolta dei dati, in sintesi, è una esigenza costosa e spesso trascurata, ma necessaria.

1.2.4 Trasformazioni del comparto produttivo

I procedimenti complessi sono storicamente diventati processi industriali in ragione dell'avvenuta parcellizzare della produzione in una sequenza ordinata di minute attività che, molto spesso, risultano così sviluppate in modo più efficiente. Da qui lo scorporo di determinati rami di produzione e la nascita di vere e proprie aziende specializzate. Specializzarsi significa, prima di tutto, delineare i contorni del proprio operare, quindi escludere tutte quelle attività che dovrebbero far parte

delle competenze specifiche di un'altra struttura produttiva (interna o esterna all'azienda). Il processo di specializzazione, fino ad arrivare addirittura a scindere funzionalmente le aziende del comparto, è favorito principalmente da due fattori: l'incremento della domanda e l'abbattimento dei costi che avviene (anche e spesso) in ragione dello stesso sviluppo indotto. Due spinte che, autoalimentandosi, consentono le concentrazioni produttive, le conseguenti economie di scala e la relativa ulteriore specializzazione. La contrazione dei costi è frutto di molteplici elementi: azioni dirette e volontarie (adozione di tecnologia, programmazione di attività ripetitive, ecc.) o indotte dalle innovazioni, dalle concentrazioni produttive e dalle relative economie di scala. Un modo per contenere i costi per un'azienda è quello di espellere i processi che si ritengono difficilmente standardizzabili (ad alta variabilità). Quel che ne rimane è un prodotto specializzato ancor più definito nel suo contorno anche se non sempre funzionale/pratico all'industrializzazione. Infatti, la specializzazione è caratteristica dell'evoluzione industriale, ma le due cose hanno sfumature diverse non di poco conto. Industrializzare significa riassemblare e coordinare attività e produzioni specializzate e presenta quindi connotati più complessi della semplice specializzazione produttiva. Deve tenere conto della natura delle attività che sono espulse dai singoli procedimenti specializzati e, quindi, della completezza del prodotto finito e dei relativi costi transazionali.

Anche la ricerca può essere vista e descritta nell'ottica di tali trasformazioni economico-produttive. La parcellizzazione dei processi nella ricerca sociale ha un impatto non solo sulle possibili modalità di acquisizione del dato e dunque sugli errori extracampionari commessi, ma anche – innanzitutto – sul comparto produttivo che ha poco o nulla a che fare con la metodologia di ricerca, ma che assume notevole importanza se si sposta l'ottica dal dato all'informazione. Non si può dimenticare, infatti, che è proprio l'informazione il prodotto finito della ricerca. L'informazione, è un elaborato concettuale che si basa, anche ma non solo, sui dati. Molti

pensano che la professionalità necessaria per elaborare/analizzare il dato ha poco o nulla a che fare con quella (strettamente legata alla conoscenza esperienziale) relativa all'elaborazione/creazione dell'informazione che dal dato può provenire. Una abilità così strettamente collegata alle capacità individuali da essere concepita come un patrimonio generico (ma non generalizzato), non necessariamente specialistico, quasi genetico e, ovviamente, afferente al decisore. Tale visione non nega la valenza del dato; al contrario, spesso la esalta fino a confonderlo con l'informazione stessa⁵⁰. Se a questo aggiungiamo l'aumentata disponibilità di dati immagazzinati in potenti *database* e la tendenza alla concentrazione di dati nel web, cioè l'opportunità offerta dal web di ampliare le platee presso cui recuperare i dati, appare evidente il motivo della crescita repentina del mercato denominato DIY (*Do It Yourself*). D'altra parte, ciò si coniuga perfettamente con la necessità di contenere i costi; soprattutto quelli sostenuti dalle società che si occupano di ricerca. Un primo e consistente taglio effettuato nel tempo dai vari istituti di ricerca è stato, per l'appunto, quello relativo ai costi di produzione dell'informazione, trasformandosi sempre più in "fabbriche di dati" ("tabellifici"). Produrre informazione, infatti, significa dotarsi di figure professionali la cui formazione è difficile e prolungata nel tempo e per questo decisamente più costose. Inoltre, la trasformazione del dato in informazione necessita di tempi adeguati, spesso non standardizzabili, che sempre più contrastano con la necessità (o velleità) di avere subito una risposta alle proprie domande. La congiunzione di questi fattori culturali e di strategia produttiva, hanno un solo difetto: rendono

⁵⁰ Il prof. Bernardi (2001), in un saggio inerente la "statistica e i mezzi di comunicazione di massa" citava il suo collega prof. Colombo che, all'atto di chiusura di un Convegno rivolto al Ministero della Pubblica Istruzione e ai presidenti degli IRSSAE, aveva detto: "Il numero non è la statistica, così come il marmo non è la scultura. Entrambi, infatti, vanno modellati per acquisire valore intrinseco ed entrambi in qualche modo godono del respiro dell'originalità e dell'estro del creatore. Ciò che differenzia lo statista dall'artista è che il primo deve fornire completa giustificazione del metodo e degli attrezzi che usa, deve assicurare cioè il principio della riproducibilità dei suoi risultati".

precaria la tesi d'essere di fronte ad una industrializzazione del processo di produzione dell'informazione⁵¹.

L'abbattimento dei costi di produzione ha favorito l'introduzione, prima, e l'ampio uso, in seguito, della ricerca socio-economica e di marketing in tutto il mondo industrializzato (incentivando la domanda di questi servizi) e ciò ha influito sul processo di specializzazione. Tuttavia, la specializzazione produttiva comporta il potenziale rischio di snaturare il fine stesso della ricerca che è quello di produrre informazione. C'è da chiedersi se abbattendo i costi di reperimento dei dati non si corra il rischio di perdere la competenza analitica. Può la competenza concettuale-elaborativa prendere vigore da una separazione netta e premeditata tra chi produce i dati e chi dovrebbe trasformarli in informazione? In questo senso ci si dovrebbe interrogare anche sulla differenza tra le università americane (e di altri paesi) e quelle italiane. Per la ricerca sociale è così saggio un modello di istruzione che separa – di fatto – la pratica di ricerca dall'insegnamento della stessa? In Italia sono quasi inesistenti i centri di ricerca sociale strettamente collegati al mondo accademico. In questo modo, secondo me, si corre il rischio di diffondere pratiche d'indagine e di generare istituti di ricerca che si confrontano essenzialmente con i fattori di produttività, senza fornire adeguati contrappesi circa la qualità del prodotto finale. E non si tratta solo di contrappesi, ma di assicurare la cultura della ricerca, ossia l'idea organica del prodotto-processo di creazione dell'informazione. La trasformazione del comparto produttivo è conseguenza e non solo artefice della modificazione della qualità dei processi utilizzati. La necessità di governare i cambiamenti non può che prendere forma nel luogo dell'istruzione, nella promozione di una cultura di ricerca basata sull'incontro tra gli elementi esperienziali e quelli strettamente teorici. In altre parole, la specializzazione produttiva non può accreditare l'idea che il percorso formativo professionale nella ricerca (sociale, così come in altre discipline) possa avvenire in un contesto di

⁵¹ A lato pratico rendono ancora più difficoltoso il lavoro del ricercatore sociale.

iperspecializzazione tecnica avulso da un organico contesto culturale. Tutto questo costituisce una sfida per le scienze sociali su cui torneremo anche più avanti.

1.2.5 Strutture centralizzate e tempi di rilevazione

Prima di passare al prossimo argomento, è forse il caso di accennare ad un ulteriore importante elemento connesso all'introduzione della tecnologia nella pratica di ricerca: i tempi di rilevazione e le strutture centralizzate. La sua trattazione, tra l'altro, tornerà utile in seguito, quando daremo conto di come si è pensata l'indagine ISSP Italia.

Generalmente si distinguono le indagini *in-line* (come il CATI e, parzialmente, anche il CAWI) da quelle *off-line* (come le indagini postali, le indagini dirette, le osservazioni, le trascrizioni). I tempi di raccolta dei dati sono decisamente diversi a seconda della tecnica d'indagine adottata e stabiliscono uno spartiacque (questo sì) consistente e sostanziale. Infatti, le indagini *in-line* hanno un loro punto di forza nella possibilità di immagazzinare i dati e le informazioni relative al processo di raccolta in tempo reale, velocizzando così sia i processi successivi (costruzione della matrice dati e sue elaborazioni) sia i necessari controlli statistici di qualità dell'intermedio processo produttivo in atto. Il CATI, in questo senso, ha rappresentato una vera e propria innovazione di processo, così come lo potrebbe diventare il WAPI e così come lo è lo stesso CAWI⁵². Da questo punto di vista l'evoluzione e la concentrazione tecnologica si è dimostrata utile e auspicabile, fatte salve le osservazioni riportate nei capitoli precedenti.

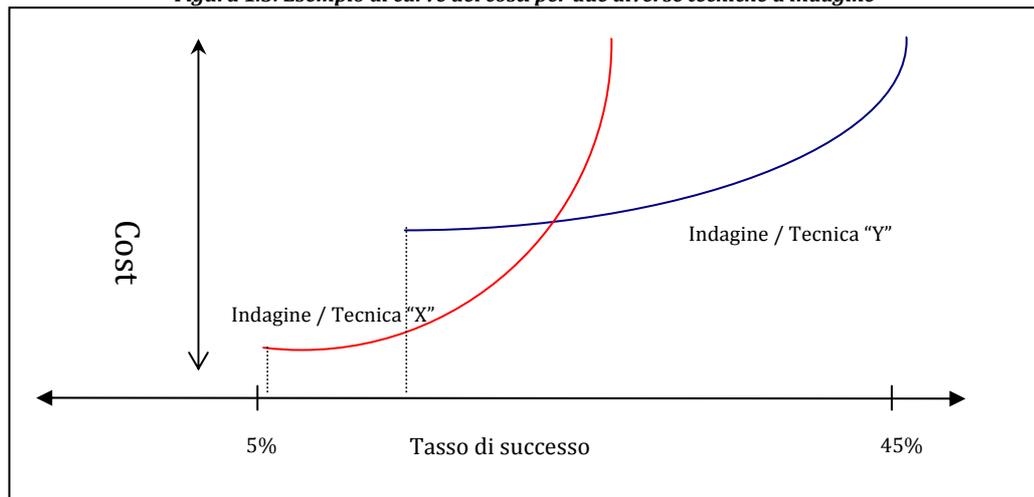
⁵² L'introduzione del computer ha comportato una drastica riduzione dei tempi di preparazione degli strumenti d'indagine (stampa dei questionari) e, ulteriore elemento importante, di quelli relativi alla validazione (test / indagini pilota).

1.3 Il valore economico della copertura campionaria: il “costo metodologico”

A parità di spesa totale, viene sempre premiato il progetto di ricerca che presenta un tasso di risposta maggiore o, meglio, il progetto che costa meno a parità di tasso di risposta. Ma come paragonare due progetti che presentano costi e livelli di adesione alquanto diversi?

Definire una funzione di “costo metodologico”, ossia attribuire un valore economico al tasso di successo, è certamente un modo per mettere a confronto piani di rilevazione completamente diversi tra loro. Si tratta, in fondo, di definire un riferimento analitico a quello che normalmente viene citato come rapporto tra qualità e prezzo.

Figura 1.3: Esempio di curve dei costi per due diverse tecniche d'indagine



Nella Figura 1.3 sono simulate due tecniche d'indagine. La prima, indagine X, mostra costi che si innalzano repentinamente all'aumentare del tasso di successo. La seconda, al contrario, evidenzia esiti positivi più elevati fin dall'inizio, ma con costi di impianto superiori e successivamente crescenti in modo meno vistoso. Sino a che le due curve non si sormontano, la tecnica X risulta vantaggiosa rispetto alla tecnica Y, ovviamente nell'ipotesi che entrambe si riferiscano alla medesima quantità d'interviste complete/effettuate. Se non si pongono obiettivi di qualità (come il raggiungimento di un determinato tasso di successo), è sempre preferibile

la tecnica che presenta costi di contatto minori. È questo, tra gli altri, il motivo del grande successo del CATI rispetto al CAPI, PAPI, WAPI.

Lo studio dei costi metodologici dovrebbe rappresentare un patrimonio di conoscenza comune, un elemento di valutazione basilare tra quanti si occupano di ricerca. Infatti, le diverse tecniche d'indagine per loro natura non sono "buone" o "cattive", semplicemente più o meno adatte all'obiettivo che si vuole raggiungere e al "valore" dell'informazione⁵³. L'utilità di applicare un criterio di valutazione come il costo metodologico è quella di definire un parametro di riferimento per una spesa atta ad abbattere al massimo il tasso di rifiuto. I costi della ricerca sarebbero comparabili tra indagini diverse, valutabili alla luce della tecnica impiegata, nonché studiati dal punto di vista longitudinale in ragione dei cambiamenti della propensione alla risposta da parte della popolazione.

È sensazione comune, anzi un fatto, che i tassi di rifiuto siano accresciuti nel tempo⁵⁴ (Keeter, 2006; Kim 2011). Ad esempio, è evidente che la minore intrusività del contatto telefonico rispetto al contatto personale, punto di forza in passato della tecnica CATI, non sia più un elemento su cui contare al fine di perseguire un buon risultato di ricerca. Complesso è quantificare la contrazione della disponibilità delle persone a partecipare all'indagine e, soprattutto, come questa nuova situazione abbia modificato la "forma" della curva piuttosto che prodotto una semplice traslazione a sinistra della stessa. Sta di fatto che, sia per costi che per tenuta sperimentale dei piani d'indagine, le "vecchie" indagini postali o *face to face* stanno riscoprendo oggi una nuova primavera dopo che il CATI le

⁵³ Si può ritenere valida qualsiasi sperimentazione come lo è, ad esempio, il CARI (somministrazione e registrazione dell'intervista senza operatore umano). Ma se questa soluzione dovesse essere scelta al solo scopo di abbattere i costi (eliminazione completa dei costi del personale addetto alla rilevazione), allora l'acquirente dovrebbe chiedersi se su un dato statistico "regalato", ma privo di ogni garanzia (in sostanza una chiacchierata) valga la pena di riporre il proprio tempo e la propria attenzione e quanto sarebbe disposto a "scommettere" sulla validità dell'informazione che ne ricava.

⁵⁴ Non sappiamo quanto ciò sia dovuto alla credibilità attribuita alla ricerca e alla percezione di scarsa utilità dei suoi scopi o ad una semplice diminuita disponibilità legata, probabilmente, ad una percezione di invadenza nella sfera personale dell'individuo contattato.

aveva quasi del tutto soppiantate. Sarebbe certamente una sfida metodologica stimolante quella di predisporre degli studi longitudinali atti a valutare l'impatto delle prassi adottate sui costi e sul tasso di adesione. Ciò, costringerebbe i ricercatori a determinare gli indicatori di qualità monitorati nel tempo e, di conseguenza, questo farebbe da contrappeso al sempre presente interesse nel perseguire il raggiungimento di un determinata numerosità campionaria, quasi senza considerare se è stato necessario contattare "n" persone o mille volte "n". È da ricordare, inoltre, che il tasso di successo potrebbe essere (volutamente) falsato da pratiche non ortodosse. In questo caso, avere la possibilità di ampliare la gamma dei parametri di riferimento qualitativi per un'indagine assumerebbe ancor più rilievo.

Prima di tornare su quest'ultimo argomento, si propone un possibile confronto tramite il costo metodologico.

1.3.1 Il confronto tra tecniche d'indagine: una proposta

Definiamo i seguenti indicatori relativi all'indagine X:

- C_X = costo totale dell'indagine X
- N_X = numerosità del campione coinvolto nell'indagine X
- n_X = campione effettivamente ottenuto dall'indagine X.

Per ogni indagine è possibile calcolare gli usuali indicatori:

- $T_X = n_X / N_X$ = tasso di risposta effettivamente raggiunto nell'indagine X
- C_X / n_X = costo pro intervista completata
- C_X / N_X = costo pro contatto.

Inoltre, possiamo anche calcolare:

- $Z_X = C_X / T_X = N_X * C_X / n_X$ = **costo metodologico** (costo dell'indagine rapportato al tasso di successo ottenuto).

Poiché la comparazione tra due indagini non può avvenire se non in ragione di un comune obiettivo campionario raggiunto, è necessario inserire un fattore di "riporto" delle due indagini alla medesima

numerosità. Estendiamo quindi gli indici sopra definiti e indichiamo $K_{XY} = n_X/n_Y$ quale moltiplicatore di confronto tra le due indagini X e Y⁵⁵.

Consapevole di imporre una forzatura rispetto al fatto che le curve di costo metodologico non sono affatto delle semirette quanto, piuttosto, delle espressioni esponenziali, per praticità e a puro scopo esemplificativo si ipotizza di confrontare le due indagini verificando se $Z_X > K_{XY} * Z_Y$. Se ciò si verifica, allora l'indagine X è da considerarsi preferibile all'indagine Y (e viceversa).

Figura 1.4: Esempio di applicazione del metodo comparativo tramite il "costo metodologico". Anno di riferimento 2012

Tipo di rilevazione	CATI (Intervista 20')		Face to Face (comparato n1/n2)	
n1/n2 = delta obiettivo raggiunto				57.6%
n= obiettivo campionario	933	1620		933
II= Unità campionate	7585	5900		3398
TS= Tasso di successo	12.3%	27.5%		27.5%
Costo Totale	€ 11 264	€ 45 800	€	26 377
C / n = Costo pro Intervista	€ 12.07	€ 28.27	€	28.27
C / II = Costo pro Contatto	€ 1.49	€ 7.76	€	7.76
C / TS /100 = Costo Metodologico (pro punto % di successo)	€ 915.73	€ 1 668.02	€	960.66

Come risulta evidente dall'esempio riportato in Figura 1.4, si osserva che l'indagine CATI è ancora preferibile a quella *face to face*; tuttavia i costi metodologici relativi alle due indagini non sono poi così differenti. È utile ricordare che il calcolo andrebbe effettuato almeno lungo tutto il dominio comune tra le due indagini cioè nell'ambito dell'intervallo tra i tassi di adesione considerati il punto "minimale" da raggiungere e quelli "realmente ottenibili" o desiderati.

Brusati (2001), nel libro "Dall'intervista alla notizia" (Tuzzi, 2001) (nota 4 – pag.171), avvalga anch'esso la tesi che i costi di produzione di una "buona" intervista *viz à viz* ed una "buona" intervista CATI non differiscono di molto tra essi. Dato il momento storico in cui scrive l'articolo, ossia in ragione del fatto i costi telefonici e quelli inerenti gli strumenti software in uso erano superiori a quelli attuali, egli arriva ad affermare che è vero il contrario, che l'intervista telefonica ha costi di produzione superiori. Ma

⁵⁵ Si noti che $K_{XY} \neq K_{YX}$, ovvero $K_{XY} = 1 / K_{YX} = n_Y / n_X$.

come qualificare quel termine “buona” intervista? Keeter (2000) usa il termine *standard* per caratterizzare una prima indagine svolta in cinque giorni di lavoro, e *rigorous* per definirne un'altra che ha esteso la rilevazione a otto settimane. Il tutto con l'identico questionario. I tassi di adesione ottenuti? 60,6% per l'indagine *rigorous* and 36,0% per quella *standard*. Con quali costi d'inefficienza indotta nella struttura di rilevazione ha ottenuto questi risultati? I tempi dedicati all'azione di raccolta dei dati sembrano indiscutibilmente connessi alla qualità dell'indagine prodotta, così com'è indiscutibile la percezione dei ricercatori circa l'esistenza di “buoni/rigorosi” modi di operare e altri, purtroppo definiti, “standard”.

L'applicazione del metodo comparativo tramite il “costo metodologico” permette al ricercatore di tarare al meglio le proprie prassi di lavoro per massimizzare l'efficacia della spesa. Facciamo in tal senso un parallelo. Nella teoria del campionamento è evidente che l'aumento della numerosità sondata migliora l'attendibilità della stima (diminuisce l'errore campionario). Però, è altrettanto chiaro che sforzi aggiuntivi rivolti a diminuire l'errore campionario oltre una certa soglia (generalmente definita al valore 95%) comportano un incremento esponenziale della numerosità da contattare. Sforzi operativi ed economici, che da un certo punto in poi sono considerati quasi inutili. Allo stesso modo, avere indicazioni di quale sia un possibile punto di rottura tra le diverse tecniche d'indagine consente al ricercatore di perseguire determinati sforzi solo in ragione dell'essere o meno prossimo a tale soglia. Chi affronta oggi una ricerca scientifica, invece, spesso non ha la reale percezione di cosa (in termini di prassi e metodi) e quanto mettere in atto in termini di spesa/sforzi. Al di là dell'esperienza personale, quindi, il ricercatore diventa un fatalista, ovvero quanto più “corpo” estraneo alla scienza.

1.3.2 Il virus degli *spare-panel*

Torniamo per un attimo al problema della falsificazione dei tassi di successo, ossia alla possibilità che dei parametri di riferimento possano

smascherare prassi di lavoro non ortodosse. Se partiamo dal presupposto che:

1. ci sono categorie sociali che difficilmente si fanno intervistare/raggiungere (giovani, lavorativamente attivi, maschi, persone particolarmente anziane e con bassi livelli di istruzione)
2. gli appuntamenti telefonici concordati hanno pari se non minore probabilità d'essere effettivamente portati positivamente a termine rispetto ad una nuova chiamata⁵⁶
3. inseguire la disponibilità altrui (relativamente ad entrambi i punti precedenti) crea una inefficienza economica e organizzativa delle strutture di rilevazione
4. i tempi necessari per raggiungere una corretta copertura della popolazione potrebbero essere incompatibili con i tempi di rilevazione (velocità con cui i dati devono essere disponibili),

allora, fare riferimento agli *spare-panel*, ossia alle persone che avevano offerto collaborazione in una precedente indagine, non è tanto la risposta al tentativo di dare rappresentatività al campione sondato (in un tempo limitato e congruo), come qualcuno vorrebbe sostenere, quanto il puntare all'abbattimento del costo di produzione. Ciò significa non solo distruggere il presupposto di casualità insito nella teoria del campionamento, ma anche dar vita ad una vera e propria manipolazione del tasso di successo. Significa, nella sostanza, inserire un virus nella possibilità stessa che la scienza si occupi di trovare le proprie pratiche migliori. Come già è stato messo in luce in precedenza, questo fenomeno è radicato, praticamente da sempre, nelle indagini CATI⁵⁷. Ciononostante, le trattazioni accademiche non evidenziano adeguatamente quanto questa prassi sia distruttiva.

⁵⁶ Ciò spiega in parte l'innalzamento repentino dei costi rispetto al tasso di successo raggiungibile.

⁵⁷ Si veda a tal proposito quanto posto agli atti da Elio Brusati (Doxa) in occasione del Convegno organizzato a Bressanone dalla facoltà di Statistica dell'Università di Padova nel 1990, quando la tecnica CATI aveva di mosso i suoi primi passi in Italia (Fabbris, 1990).

È vero che nelle indagini (commerciali e non solo) spesso si usano campioni per quote dove la componente sperimentale è teoricamente anche del tutto assente. Ma è altrettanto vero che, almeno nei bandi di gara relativi alle indagini ufficiali promosse da istituzioni pubbliche alle quali tutti riconoscono un presunto rigore, ci si aspetterebbe il divieto assoluto di utilizzare questi *spare-panel*. Perpetua dimenticanza, scarsa sensibilità della cultura di ricerca, limitata conoscenza degli scenari del mercato della produzione? Non si tratta di uno sterile atto di accusa; più semplicemente la constatazione che perseguire come unico obiettivo il minor costo senza un serio raffronto e studio circa i risultati ottenuti nelle ricerche stesse, conduce la ricerca sociale applicata in un terreno arido e incontrollato. Paradossalmente, chi viene premiato per la capacità di ottenere elevati tassi di successo nelle indagini, dovrebbe in realtà – “grazie” agli *spare-panel* – essere additato per queste degenerazioni metodologiche e, aggiungerei, per turbativa di mercato, avendo la possibilità di abbattere oltremodo i costi (Brusati, 2001)⁵⁸. Infatti, se si svolgono diverse indagini presso lo stesso nucleo familiare a distanza di qualche mese l’una dall’altra, l’indice di concordanza grezzo tra gli esiti del contatto (intervista completa, rifiuto, occupato, libero, appuntamento mai rispettato, ecc.) è pari a circa il 24%-25%⁵⁹. Significa che circa un quarto delle famiglie adotta la medesima reazione ogni volta che viene sollecitata a partecipare ad un’indagine. Ed ancora, se ci si limita ad analizzare i tentativi di contatto con le famiglie che hanno aderito la prima volta ad un’indagine, l’indice di concordanza sale a circa il 30%. In altre parole, una famiglia su tre si rende disponibile a partecipare ad un’ulteriore indagine. Il tasso di adesione che generalmente

⁵⁸ Scrive l’autore: “Perciò, chi vende a troppo basso prezzo i sondaggi telefonici, battendo troppo facilmente tutti i concorrenti, può farlo solo o perché non gli importa di rimetterci (per esempio è un mecenate, o ha altri fini), o perché bara, E ci sono mille modi per barare:...”.

⁵⁹ I dati presentati sono elaborazioni statistiche svolte dalla società Marker Srl di Mestre-Venezia su oltre un milione di contatti CATI. In molti casi la ripetizione della medesima indagine (ad esempio le prospettive di voto su piccoli-medio comuni) comportava la necessità di riutilizzare tutti i numeri telefonici presenti in un’area. Aggiungendo a questi casi anche quelli per i quali la casualità di estrazione aveva proposto lo stesso numero di telefono, si sono ricavati i dati esposti. I dati non compaiono in nessun articolo o rivista essendo analisi interne alla società.

si raggiunge oggi presso la popolazione complessiva nelle indagini CATI varia (a seconda dei tempi complessivi dell'indagine, dei temi trattati e delle modalità di contatto) mediamente dal 5% al 15%. Raramente si arriva al 20%. Va da sé che l'uso degli *spare-panel*, cioè l'utilizzo dei numeri telefonici con risposta positiva in precedenti sondaggi, comporta un vantaggio competitivo (risparmio) che può arrivare anche al 50% della spesa. A questo tipo di prassi non si pone facilmente rimedio, ma se ci fosse la possibilità di comparare le qualità delle metodiche d'indagine attraverso appropriati indicatori, questi "miracoli della fantasia umana" potrebbero essere maggiormente smascherati/monitorati.

1.4 Sfide alle scienze sociali

Partendo dai costi si è cercato di affrontare un tema che interessa l'insieme delle trasformazioni ambientali (economiche e sociali), dove le tecnologie rappresentano una risposta, nonché un elemento propulsore a un cambiamento generale che da sempre coinvolge l'individuo e la società nel suo complesso. La scienza della ricerca sociale non è estranea a questa trasformazione.

In una visione generale, distaccata e lontana dalle realtà temporali e territoriali, si è portati a valutare la storia dell'uomo come un continuo percorso in progress e quasi mai come un cammino fatto di tentativi, di regressioni e cambiamenti circostanziati nel tempo e nello spazio. Invece, le idee e i rinnovamenti che scandiscono la storia non sempre hanno percorso la via del progresso. L'innovazione, tanto cara a Schumpeter (1932, 1939) quale meccanismo di accrescimento economico e quindi anche sociale, assume un carattere positivo quando è frutto di un percorso di trasformazione razionale e condiviso e non quando le trasformazioni (che non sono propriamente innovazione) assumono i caratteri di una deriva dei processi che prendono vigore dalle trasformazioni tecnologiche, competitive e "ambientali", come alcune di quelle descritte nei capitoli precedenti. Puntare l'attenzione sui processi produttivi della ricerca sociale è il tentativo di razionalizzare il panorama che ci circonda per comprendere se corrisponde almeno in parte all'orizzonte verso cui si vuole tendere. In fondo, la strada la scegliamo sempre e comunque noi!⁶⁰

⁶⁰ Se invece di parlare di ricerca sociale si trattasse della produzione agricola ed alimentare, il ragionamento diverrebbe più chiaro e decisamente più sentito. L'uso indiscriminato della chimica è stata vissuta per molti anni come un'innovazione, capace di incrementare le scorte alimentari e alleviare il lavoro necessario nella produzione. L'abuso da un lato, ma anche l'insieme dei costi (diretti ed indiretti) che ingenera dall'altro, porta oggi a rivedere l'atteggiamento complessivo verso l'impiego di prodotti chimici e delle colture geneticamente modificate. Non si tratta di mistificare o di avere atteggiamenti censori, ma semplicemente di prendere atto che qualsiasi percorso intrapreso porta con sé delle conseguenze e delinea degli scenari futuri non sempre ignoti, ma molto più spesso inevitabilmente noti, prevedibili, accettati.

Il dibattito teorico sui processi evolutivi economici, quasi del tutto estranei alla teoria classica, individua nell'autopropulsione⁶¹, nella tendenza alla concentrazione⁶² e nell'innovazione⁶³ gli elementi genetici di cambiamento strutturale degli scenari economici e sociali. Ciò che noi identifichiamo come scienza prende necessariamente spunto dalla storia e dovrebbe essere la risposta metodologicamente corretta alle problematiche del tempo (Rao, 2008). Ma se le trasformazioni economiche e sociali sono frutto dei mutamenti organizzativi e produttivi, allora ci si chiede come queste influenzino l'evoluzione della scienza. Sarebbe anche interessante capire se le categorie dei problemi affrontati dalla ricerca sociale (in generale e in quella scientifica in particolare) e gli strumenti che essa adotta non si accreditino l'un l'altro o, meglio, se le specializzazioni produttive (che tendono ad espellere sia alcune figure professionali sia alcune parti costituenti la qualità del dato), la velocità con cui si vuole disporre dei risultati e la confusione esistente tra il concetto di dato e quello di informazione, ossia tutto ciò che comporta il non governo nella parcellizzazione delle attività di produzione della ricerca, rappresentino dei "caratteri ambientali" – dei veri e propri elementi di pressione – che possono riflettersi nel medio periodo sulla capacità di produrre risultati concreti nella ricerca scientifica e, a lungo andare, sui percorsi cognitivi da applicare per salvaguardare la scientificità stessa della ricerca.

⁶¹ Arthur Young (1928) e successivamente Nicholas Kaldor (1972) sostenevano che le trasformazioni produttive autoalimentano ulteriori cambiamenti, creando così una spirale autopropulsiva fonte dei processi di crescita o di decrescita (economie o diseconomie). Il processo cumulativo di suddivisione del lavoro, quindi, immette i semi per le trasformazioni economiche che si possono verificare nel tempo.

⁶² Piero Sraffa (1928) nella sua critica alla teoria economica liberale paventava l'idea che le concentrazioni produttive fossero fonti inequivocabili delle economie di scala e queste, a loro volta, consentissero la possibilità di investimenti ulteriori sui processi di produzione. Il suo intento volgeva all'idea che la pianificazione produttiva controllata potesse portare ad uno sviluppo economico più incisivo rispetto a quanto espresso dall'opinione liberale che individuava nella concorrenza l'interesse generale della società.

⁶³ Joseph Alois Schumpeter (1932, 1939), economista liberale, riconosceva nell'innovazione i punti cardine dei cicli economici che si potevano leggere nella storia. Che l'innovazione fosse frutto di una nuova inaspettata scoperta o il risultato di un percorso razionale/scientifico, poco importava. Egli sosteneva che la storia stessa, ovvero la demografia e la società che si va a conformare, fosse frutto dell'evoluzione economica, cioè dell'innovazione dei processi e dei prodotti che si "materializzano" nella storia. La ciclicità economica (a quattro fasi) era il frutto dell'espansione nel tempo delle innovazioni prodotte.

Le trasformazioni, in questo senso, potrebbero diventare vere e proprie derive o, quantomeno, rifrazioni culturali. Lo diventerebbero quand'esse premiassero, ad esempio, l'idea che l'analisi del contingente, la quantificazione degli umori presenti nella società, il sondaggio d'opinione e la sua necessità d'essere pronto consumo sia il percorso da intraprendere per isolare i temi da affrontare e le sfide contemporanee circa l'evoluzione della società.

La mancanza di voci dissonanti rispetto all'introduzione massiccia del CAWI su *spare-panel* quale migliore metodologia per disporre di un risultato immediato, il contrabbandare per ricerca scientifica l'analisi dei fenomeni emergenti (e sempre diversi tra loro) sui *social network*⁶⁴, cioè il tentativo di piegare la ricerca sociale all'uso di strumenti utili (forse) a precise necessità tipiche del marketing, potrebbero rappresentare causa ed effetto di uno scarso interesse attribuito ai rilevanti e attuali temi che le scienze sociali dovrebbero affrontare. Assumerebbe allora una nuova prospettiva, quantomeno in Italia, la difficoltà di trovare campioni probabilistici e/o relative discussioni su come affrontare la questione della copertura campionaria, lo scarso finanziamento alla ricerca e l'insufficiente attenzione data a quei pochi istituti che di essa si occupano.

Su questi temi prospettici sarebbe il caso che la comunità scientifica aprisse un confronto. Per questo l'Italia farebbe bene a rimanere all'interno di gruppi internazionali di ricerca che si dovrebbero interrogare sui temi sociali che permeano la nostra – così vicina – “grande società allargata”. In questo senso, quando si parla di qualità non ci si può esimere dall'affrontare anche il dibattito sulle categorie tematiche oggetto d'osservazione.

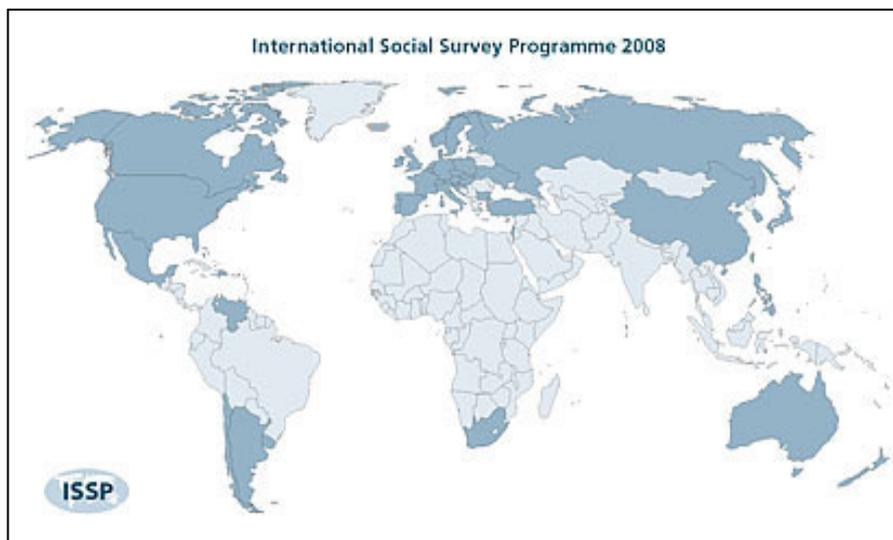
⁶⁴ Fenomeno la cui portata sarà visibile nel breve periodo e su cui i proprietari dei *social network* stanno investendo capitali enormi.

2 – L'indagine scientifica ISSP

2.1 ISSP: breve presentazione

L'International Social Survey Programme (ISSP) è un'organizzazione presente in 48 paesi. Attiva dal 1985, promuove programmi di ricerca scientifica comparativa nel tempo e tra nazioni. Tra gli istituti di ricerca che hanno dato origine all'organizzazione ricordiamo: il *National Opinion Research Center* (NORC) dell'Università of Chicago (promotore dell'indagine GSS - *General Social Survey* negli Stati Uniti); il *Zentrum für Umfragen, Methoden und Analysen* (ZUMA) in Germania (Mannheim) e il *Social and Community Planning Research* (SCPR) di Londra.

Figura 2.1: Paesi membri di ISSP (situazione al 2008)



Dopo una consolidata collaborazione, congiuntamente alla *Research School of Social Sciences (Australian National University)*, questi istituti si sono ripromessi di sviluppare e promuovere, presso altri paesi, dei moduli di ricerca comuni nell'ambito di importanti aree tematiche delle scienze sociali⁶⁵. L'Italia nell'ISSP è stata in questi ultimi anni rappresentata dall'Istituto di Ricerca Sociale dell'Università del Piemonte Orientale "Amedeo Avogadro" di Alessandria⁶⁶. Altri istituti universitari, come il

⁶⁵ Per ulteriori informazioni sull'organizzazione, i membri e le indagini sviluppate, si faccia riferimento al sito internet <http://www.issp.org>.

⁶⁶ Con responsabile scientifica la prof.ssa Cinzia Meraviglia.

Dipartimento di Scienze Sociali e Politiche dell'Università di Milano⁶⁷, avevano iniziato a collaborare a questo programma, mentre il mio personale contribuì a questa iniziativa di ricerca inizia alla fine del 2010 con lo sviluppo del piano operativo e di campionamento adottati a partire dal 2011⁶⁸.

L'adesione dei vari membri all'organizzazione, così come la metodologia utilizzata in ogni paese, è controllata da un Comitato Scientifico Internazionale (il *Methodology Committee*). Attento in primo luogo al piano di campionamento, al Comitato spetta il compito di accettare o bandire il disegno metodologico complessivo che ogni nazione intende applicare.

I moduli di ricerca che ISSP possono anche coesistere con altre indagini nazionali⁶⁹, ma il tutto deve essere svolto in modo tale da garantire la scientificità della rilevazione e quindi la comparabilità dei dati tra nazioni. In tal senso, ISSP pone dei vincoli che, relativamente agli argomenti qui trattati, vedremo nel dettaglio in seguito (cfr. cap. 2.2). Non è oggetto di tesi la questione della comparabilità semantica dei moduli d'indagine e dell'intenso lavoro necessario affinché i risultati siano tra loro confrontabili nel tempo e tra nazioni. È un argomento vastissimo se si pensa che: «[...] ISSP non si configura come un "agglomerato" di indagini armonizzate *ex post*» (Gauthier, 2000), ma come un'unica ricerca «[...] disegnata e destinata per essere realizzata in diversi gruppi culturali» (Harkness, 2003). «Il disegno di ricerca delle indagini ISSP è consapevolmente orientato alla comparazione scientifica dei dati tra paesi e nel tempo, poiché i suoi moduli tematici vengono ripetuti nel corso degli anni. Ciò offre

⁶⁷ Con referente il prof. Maria Antonio Chiesi.

⁶⁸ La mia partecipazione inizia verso la fine della prima rilevazione svolta nel 2010 dall'Università del Piemonte Orientale. Scopo della mia presenza era quello di rivedere le prassi di lavoro e il metodo di campionamento inizialmente adottati dalla società incaricata della rilevazione. La prima esperienza, infatti, aveva manifestato gravi ritardi nella raccolta dei dati, nonché insufficienti garanzie circa l'aderenza al piano di campionamento sperimentale inizialmente pensato.

⁶⁹ I moduli di ricerca possono essere inseriti all'interno di altri questionari somministrati alla popolazione. Il NORC, ad esempio, raccoglie i dati ISSP attraverso la *General Social Survey*.

un vantaggio sostanziale per l'analisi del cambiamento sociale, poiché la prospettiva cross-culturale si coniuga alla prospettiva storica, consentendo analisi di grande profondità e al tempo stesso corrette dal punto di vista metodologico» (Bianco, 2009). Le analisi comparative transnazionali assumono sempre maggiore importanza se si pensa alla necessità di dover governare delle trasformazioni che avvengono, oggi più che mai, a livello globale. A testimoniare la valenza e l'interesse verso questa iniziativa di ricerca lo è il continuo incremento delle pubblicazioni e degli studi condotti proprio a partire dai dati raccolti. Dati, tra l'altro, che sono a disposizione di tutti gratuitamente nel sito GESIS (*Leibniz Institute for Social Sciences*) all'indirizzo <http://www.gesis.org/en/issp/data-access/>.

Figura 2.2: indagini ISSP - Quantità di pubblicazioni - Anni 2002-2012

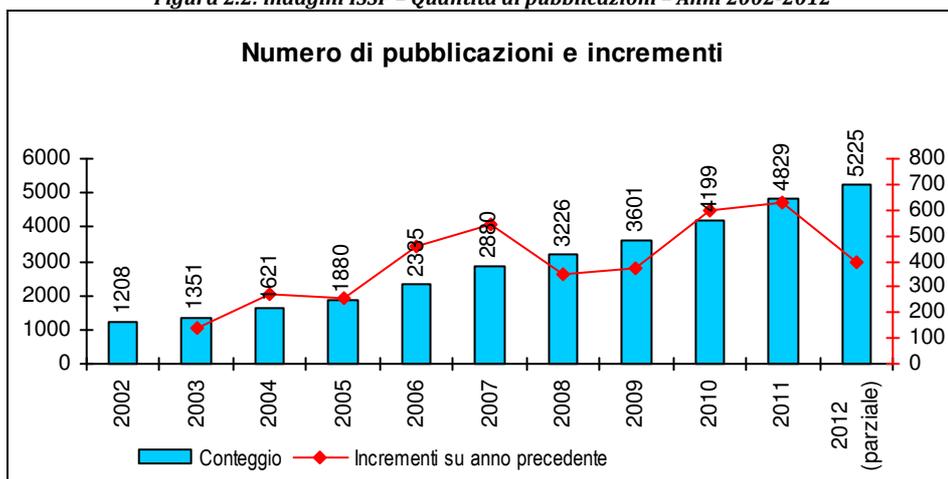
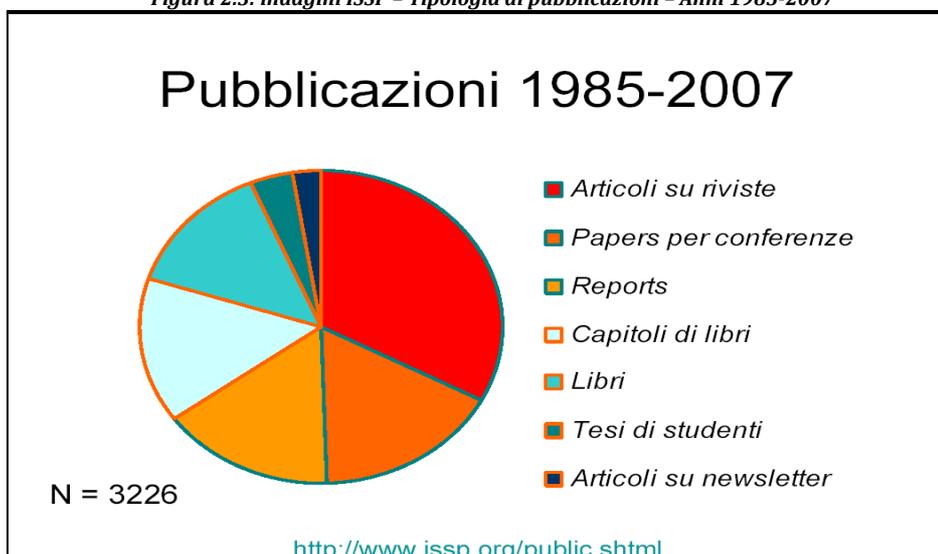


Figura 2.3: indagini ISSP - Tipologia di pubblicazioni - Anni 1985-2007



Per completare la panoramica generale, vediamo quali sono gli argomenti trattati nelle indagini, come avviene la loro scelta e quali sono le fonti di finanziamento. Ogni anno è programmata un'indagine su temi individuati collegialmente. I paesi membri non sono obbligati a dar corso all'indagine, ma la loro assenza per più annualità li esclude automaticamente dall'organizzazione. È per questo motivo che l'Italia non ha avuto per lungo tempo un rappresentante nell'organizzazione internazionale e, se non darà corso alle rilevazioni nel 2014, uscirà nuovamente dal gruppo. I temi e la programmazione delle attività di ricerca sono illustrati nella Figura 2.4. L'Italia dopo essere rientrata nell'organizzazione internazionale di ricerca nel 2009, ha fin da subito avviato in ritardo il crono programma. Nel 2011 si è sondato l'argomento delle «disuguaglianze sociali», che a livello mondiale era previsto nel 2009. Nel corso del 2012, attraverso l'adozione di un unico questionario che riuniva due moduli, sono stati sondati i temi «salute» e «ambiente». Se ISSP Italia troverà i fondi necessari, nel 2014 si affronteranno i moduli «identità nazionale», «cittadinanza» e «famiglia e cambiamento dei ruoli di genere». Si sono immaginati 3 strumenti di rilevazione distinti (questionari) che combinano 2 temi ciascuno. In tal modo, oltre ad ottenere delle economie sulla spesa (com'è avvenuto nel 2012), l'Italia potrebbe mettersi al passo con gli altri paesi.

Figura 2.4: Temi trattati nelle indagini ISSP e loro programmazione

Tema d'indagine	Anno di programmazione (a livello internazionale)
Role of Government	1985, 1990, 1996, 2006 , 2016
Social Networks	1986
Social Inequality	1887, 1992, 1999 , 2009
Family and changing gender roles	1988, 1994, 2002 , 2012
Work Orientations	1989, 1997, 2005 , 2015
Religion	1991, 1998, 2008
Environment	1993, 2000 , 2010
National Identity	1995, 2003 , 2013
Social relation and support systems	2001,
Citizenship	2004 , 2014
Leisure Time and Sports	2007 ,
Health	2011

Nota: Gli anni barrati sono quelli nei quali l'Italia non ha partecipato alla rilevazione. Quelli incorniciati riguardano le rilevazioni già programmate alle quali l'Italia, con ampia probabilità, non prenderà parte.

Ogni membro che partecipa a ISSP si incarica di trovare i fondi necessari per svolgere le proprie indagini. In alcuni paesi lo stato provvede ad un finanziamento costante, mentre in altri (come l'Italia) i ricercatori devono individuare i propri finanziatori. Se da un lato questo aspetto pone dei consistenti limiti circa la continuità stessa della ricerca, dall'altro assicura la piena autonomia dalla politica e da altri eventuali organi dello Stato. In sostanza, i membri dell'organizzazione agiscono solo in conformità alle rigorose procedure e standard di qualità decise da ISSP e valide per tutti i paesi che vi partecipano. In Italia il progetto ha partecipato a dei bandi di ricerca ricevendo parte dei finanziamenti necessari dal MIUR e, prima ancora, da alcune fondazioni bancarie.

2.2 Vincoli del comitato scientifico ISSP sull'impianto dell'indagine

Il Comitato Scientifico di ISSP pone diversi vincoli ai membri aderenti. Tra quelli relativi al disegno d'indagine, alcuni fanno riferimento al piano di campionamento, altri alla metodologia di somministrazione del questionario. Relativamente al **piano di campionamento**, è stabilito che:

- sia di tipo sperimentale/probabilistico
- nessuna sostituzione delle unità campionate, per "prossimità" con altre unità appartenenti all'universo di riferimento, deve essere prevista e praticata
- è ammesso il *continuous sampling*, ovvero la ripetizione del medesimo esperimento di selezione al fine di raggiungere l'obiettivo numerico previsto, una soglia minima di 1.000 interviste utili per ogni nazione.

I due pilastri della ricerca scientifica campionaria, ossia l'esistenza di un esperimento casuale (probabilistico) nella scelta dei soggetti da intervistare e la (potenziale o reale) completa raggiungibilità di tutto l'universo di riferimento, devono essere specificatamente salvaguardati. Non è ammesso, in nessun caso, che un'unità sia sostituita da un'altra in quanto "prossima", ovvero avente caratteristiche ritenute simili a quella estratta in prima battuta; sia che questa sia messa a disposizione dal gruppo di ricerca (definita a priori senza un metodo di campionamento sperimentale), sia che venga rintracciata dal rilevatore sul campo in base a caratteristiche specifiche valutate allo scopo. Un soggetto che non si riesce ad intervistare, dunque, va "abbandonato/fatto decadere" e, semmai, una nuova estrazione casuale dovrà selezionare un'ulteriore unità campionaria. Questa scelta, che dal punto di vista teorico e sostanziale potrebbe anche

essere opinabile⁷⁰, assume notevole rilevanza dal punto di vista pratico. Le indagini ISSP, infatti, sono generalmente svolte contattando direttamente la popolazione presso le proprie abitazioni. La prossimità a cui si fa riferimento, di conseguenza, acquisisce anche un carattere “spaziale” nel momento in cui il rilevatore, errando, sarebbe spinto a sostituire l’unità campionaria con un qualsiasi altro vicino di casa (propria o dell’individuo che doveva contattare). Il fatto che sia il ricercatore stesso a dover svolgere un’ulteriore estrazione tutela dalla tentazione di sostituire a casaccio (e non casualmente) le unità individuate in prima battuta. Null’altro è imposto dal punto di vista del piano di campionamento: non è promosso un campionamento casuale semplice, in blocco, stratificato o qualsivoglia altra strategia tra i campionamenti probabilistici possibili.

Relativamente alla **tecnica di somministrazione del questionario**, è decisamente gradita l’autosomministrazione, ma è altrettanto accettata l’intervista diretta con rilevatore, purché il tutto avvenga con il supporto di un questionario cartaceo o informatizzato che consenta all’intervistato di prendere visione degli eventuali elementi grafici usati. Ciò lascia spazio alle seguenti modalità di indagine:

- Postale (autosomministrata)
- Diretta/personale (come il [P-C-W].A.P.I. – *Paper, Computer or Web Assisted Personal Interviewing*⁷¹)
- C.A.W.I. (*Computer Assisted Web Interviewing*) (autosomministrazione via internet).

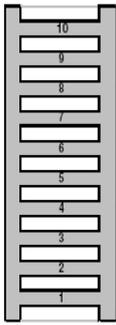
⁷⁰ Molti sono infatti gli studiosi contrari alla possibilità di ricampionare per poter raggiungere una numerosità desiderata (Booth, 1999). Infatti, se i motivi di non adesione all’indagine fossero in qualche modo correlati ai temi trattati, continuare a ricampionare significherebbe perpetuare una strategia che porta inevitabilmente ad una selezione di soggetti “distorti” rispetto alla realtà che si vuole rappresentare, un gruppo di persone avente un particolare atteggiamento rispetto al tema trattato nell’indagine stessa.

⁷¹ Grazie alla diffusione delle connessioni dati *wireless* oggi presenti sul territorio, nulla impedisce che l’intervista possa avvenire con la trasmissione in tempo reale dei dati raccolti da un rilevatore a tu per tu con l’intervistato.

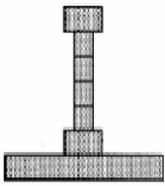
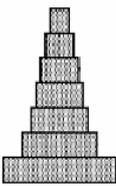
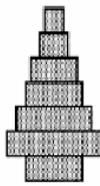
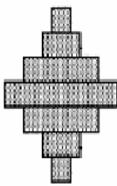
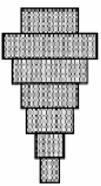
I contenuti presenti nei moduli proposti da ISSP, a volte si avvalgono di elementi grafici ritenuti essenziali affinché l'intervistato possa individuare la propria risposta e, in generale, sono concepiti per una risposta meditata⁷².

Figura 2.5: Esempi di domande con ausilio grafico

B 12. Nella nostra società ci sono persone/gruppi di persone che occupano posizioni elevate e altri che occupano posizioni più basse. Lei personalmente, dove si collocherebbe sulla scala riprodotta su questo cartellino? (sceglia una sola risposta)

	Posizioni elevate	<input type="checkbox"/> 10	Posizioni elevate
		<input type="checkbox"/> 9	
		<input type="checkbox"/> 8	
		<input type="checkbox"/> 7	
		<input type="checkbox"/> 6	
		<input type="checkbox"/> 5	
		<input type="checkbox"/> 4	
		<input type="checkbox"/> 3	
		<input type="checkbox"/> 2	
	Posizioni basse	<input type="checkbox"/> 1	Posizioni basse

B 17. Le cinque figure disegnate su questo cartellino descrivono diversi tipi di società. Dopo aver letto le descrizioni e guardato le figure mi dica...

				
Tipo A Una piccola élite in cima, pochissime persone nel mezzo e la gran massa delle persone in fondo.	Tipo B Una società come una piramide con una piccola élite in cima, più persone nel mezzo e la maggior parte in fondo.	Tipo C Una piramide, a eccezione del fatto che solo poche persone sono in fondo.	Tipo D Una società con la maggioranza delle persone nel mezzo.	Tipo E Molte persone vicine alla cima e solo poche persone vicine al fondo.

Con tali vincoli, è chiaro il motivo per cui la tecnica CATI non è ammessa: non risponde alle specifiche circa la possibilità d'utilizzo di eventuali supporti grafici e, per sua natura, cioè per il fatto che è una forma di raccolta dati che si fonda sul solo confronto verbale tra intervistato e intervistatore, risente di una velocità d'esecuzione spesso nemica di quella risposta meditata che si desidera ottenere. Inoltre, i vincoli di campionamento posti non possono essere rispettati usando il CATI, o meglio la lista telefonica. Il CAWI, al contrario, è accettato sia perché

⁷² Contrariamente a quello che accade nelle indagini di marketing, nei moduli d'indagine ISSP non sono mai presenti domande che necessitano di un tempo limitato di risposta; così come lo potrebbe prevedere, ad esempio, la seguente domanda: "Mi può dire la prima cosa che le viene in mente (potrebbero essere marche, aggettivi, ecc.) quando si parla di?".

implementa un supporto grafico, sia per il fatto che asseconda i tempi che ogni intervistato dedica a ogni singola risposta.

2.3 Ricognizione operativa per la predisposizione dell'indagine in Italia

Lo scenario del mercato della ricerca e le prassi da adottare sono state un tema non indifferente nella preparazione dell'indagine ISSP in Italia. Si è cercato di rendere esplicito l'impegno economico con il committente, soppesando l'utilità metodologica delle prassi da adottare in ragione anche dei costi connessi alla salvaguardia della qualità desiderata. Normalmente si parla – minimizzando forse la portata del tema – di analisi del rapporto tra qualità e prezzo.

Prima di illustrare la soluzione adottata è utile accennare alla disponibilità di altre informazioni già presenti nel momento in cui si sono definite le scelte: l'indagine pilota e la possibilità di affiancare ISSP ad altre indagini già in corso.

2.3.1 L'indagine pilota

Nel 2010 l'Istituto di Ricerca Sociale dell'Università del Piemonte Orientale ha predisposto un'indagine postale su 300 famiglie/individui scelte/i casualmente tra gli indirizzi dell'elenco telefonico. Lo scopo non era tanto quello di predisporre una minuziosa indagine scientifica, quanto di "saggiare" quali potessero essere i tassi di adesione, simulando e valutando l'impatto delle procedure pensate all'epoca: l'invio di una lettera di presentazione dell'indagine; la spedizione del questionario; le lettere di sollecito alla compilazione e restituzione del questionario, ecc. All'interno della famiglia, un solo individuo era tenuto a rispondere al questionario. Il soggetto era colui/colei che aveva la data di nascita più vicina ad una data casuale (del tipo giorno-mese) generata per ogni questionario. Il tasso di risposta finale, a fronte di svariati solleciti (sempre postali) e dell'invio di

un ulteriore questionario, è stato del 17-18%, comprensivo anche dei pochi questionari restituiti parzialmente compilati.

2.3.2 La disponibilità di campioni probabilistici in Italia

Potendo l'indagine ISSP coesistere con altre indagini, in via del tutto preliminare i responsabili della ricerca in Italia hanno approfondito questa possibilità valutando, nello specifico, i disegni di ricerca e i piani di campionamento che noti e accreditati istituti utilizzano per altre indagini. Lo scenario che si è prefigurato è stato pressoché questo:

- rari sono gli istituti di ricerca in Italia che adottano (o hanno la possibilità di adottare) campionamenti completamente probabilistici presso la popolazione italiana
- basati su liste elettorali, i piani operativi proposti fanno spesso sorgere il sospetto che i punti di campionamento siano scelti in funzione della disponibilità di una rete di rilevatori sul campo, piuttosto che il contrario
- l'onerosa gestione delle indagini *face to face* presso le famiglie spesso evidenzia carenze dal punto di vista del controllo della rete dei rilevatori e risulta elevato il tasso di sostituzione delle unità e l'utilizzo di criteri di prossimità/vicinanza spaziale per la sostituzione degli individui estratti in prima battuta.

In sintesi, la possibilità di affiancare l'indagine ISSP a qualche altra rilevazione è, per quanto riguarda il panorama degli istituti privati di ricerca in Italia, pressoché nulla. La strada perseguita dai responsabili della ricerca, di conseguenza, è stata quella di creare un nuovo e proprio progetto di rilevazione⁷³.

2.3.3 Schema sintetico dei principali ostacoli e obiettivi

Si sono dovute affrontare e risolvere le seguenti problematiche:

⁷³ Decisione da cui poi è derivata la mia collaborazione.

➤ Metodo d'indagine

- La tecnica CATI è esclusa a priori da ISSP
- L'indagine personale ha costi (diretti e indotti) elevati

➤ Piano di campionamento

- Data la scarsa copertura della popolazione residente, l'elenco telefonico non è utilizzabile neppure quale fonte per la definizione di un esperimento casuale di scelta delle unità di campionamento
- Le liste anagrafiche ufficiali (anagrafi comunali ed elettorali) attraverso le quali costruire un'indagine postale:
 - non sono reperibili in tempi certi e contenuti
 - necessitano di autorizzazioni alla consultazione.

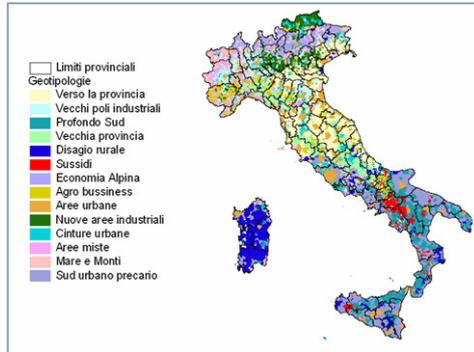
Per minimizzare il costo metodologico, salvaguardando allo stesso tempo i vincoli metodologici imposti, si è cercato di mettere in campo nuove idee al fine di:

1. contenere i costi diretti e indotti
2. preservare la qualità dell'indagine con processi di controllo più simili a quelli adottati nelle indagini "in-line" piuttosto che "off-line"
3. individuare un metodo (o delle liste) tale da ottenere la massima copertura dell'universo campionario
4. abbattere la soglia di diffidenza e disinteresse della popolazione verso la ricerca.

2.4 Il piano di ricerca ISSP in Italia: cenni della soluzione adottata

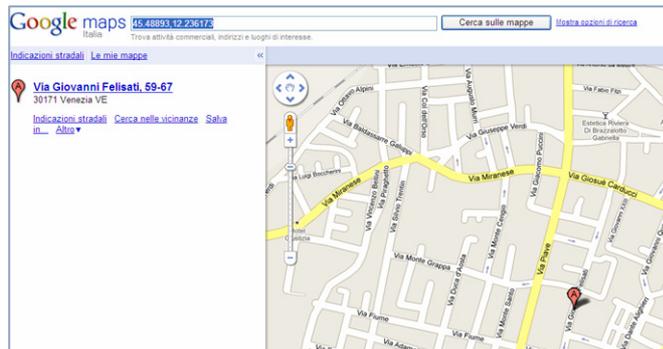
2.4.1 Il piano di campionamento

Per l'indagine ISSP Italia si è formulato un **piano di campionamento** sperimentale (a quattro stadi autoponderato), che:

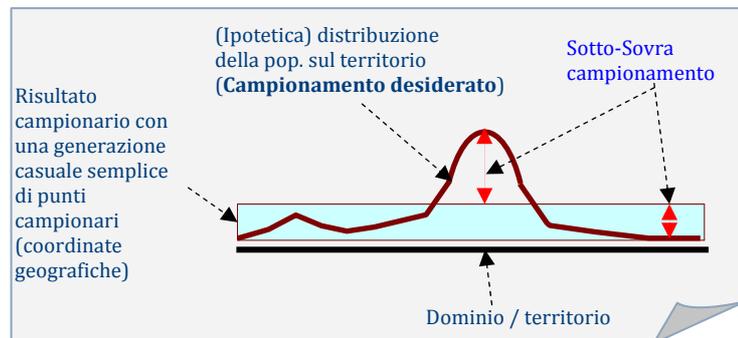


- stratifica il territorio in base alle omogeneità socio-economiche dello stesso (geotipologie),

- sfrutta le informazioni di navigazione GPS (campione geocasualizzato),



- rispetta lo “sviluppo verticale” delle realtà urbane, ossia tiene conto della densità abitativa all’interno dell’area comunale,



tutti elementi di novità e attualità nel panorama internazionale della ricerca sociale. In ragione di determinate peculiarità delle informazioni di base utilizzate, in Italia si è seguito un percorso diverso ed autonomo

rispetto a quello che successivamente si è scoperto essere stato proposto dal Dipartimento di Geografia dell'Università dell'Iowa per un'indagine pilota nell'area metropolitana di Chicago negli Stati Uniti (Kumar et al., 2011). Come si evidenzierà in seguito, i due percorsi di ricerca, senza saperlo, sono sostanzialmente giunti alla medesima proposta pur partendo da condizioni iniziali diverse. Ciò rende il metodo di campionamento ancora più interessante, data l'elevata capacità di adeguarsi a diverse situazioni/condizioni iniziali.

2.4.2 Il piano operativo di rilevazione

Relativamente alla **tecnica di somministrazione**, si è deciso che:

- L'indagine si sarebbe svolta attraverso il recapito personale alla famiglia di un questionario cartaceo da parte di un rilevatore selezionato ad hoc.
- L'intervistato/a (scelto/a casualmente tra i membri della famiglia attraverso il metodo del confronto tra una data casuale e quella del proprio compleanno) avrebbe avuto la possibilità di rispondere:
 - via CAWI
 - via posta, tramite busta pre-affrancata.
- Nessun incentivo, economico o materiale, sarebbe stato fornito all'intervistato quale compenso per la sua partecipazione.

L'**organizzazione** si sarebbe appoggiata ad una struttura di coordinamento ad hoc con il compito di:

- predisporre i materiali necessari (amministrativi, informatici e cartacei da far recapitare alle famiglie)
- implementare il campione sperimentale
- selezionare e istruire i rilevatori
- offrire supporto (informativo ed eventualmente tecnico) agli intervistati e ai rilevatori

2.4.3 Una possibile definizione dell'indagine ISSP in Italia

Da quanto brevemente descritto e schematizzato nella Figura 2.6, si potrebbe dire che l'indagine ISSP Italia è un'**indagine diretta (ma non personale), svolta con tecnica ibrida ("semi-postale" e CAWI) su un campione sperimentale geo-casualizzato della popolazione italiana**. Il disegno d'indagine è stato progettato cercando di recuperare, da varie esperienze di ricerca, gli aspetti metodologicamente "migliori" dal punto di vista del rapporto qualità-prezzo. Quali esempi, si può citare il fatto che si è preferito:

- tentare una consegna diretta del questionario, invece di raccogliere il solo riferimento della famiglia e utilizzare poi il tramite postale
- far riversare le informazioni dell'avvenuto recapito su moduli web piuttosto che attendere la consegna alla struttura centrale delle schede cartacee compilate dai rilevatori
- effettuare i solleciti centralmente tramite struttura CATI e non far tornare il rilevatore presso la famiglia
- mettere a disposizione almeno due modalità di compilazione/trasmissione del questionario non vincolando al solo metodo postale.

Tuttavia, non è tanto l'organizzazione e le prassi adottate ad essere una novità ed un punto rilevante della ricerca, quanto il metodo di campionamento che sarà in larga misura il tema dei capitoli successivi.

Uno sguardo al comparto produttivo, considerazioni sul processo di creazione dell'informazione, così come una valutazione delle opportunità offerte dalle tecnologie emergenti, è sembrato l'approccio più consono al fine di proporre un piano di ricerca che fosse coerente con quanto richiesto dal Comitato Scientifico internazionale ISSP. Si è cercato di coniugare la sostenibilità economica con quella metodologica. Anche in questo senso vanno lette le varie riflessioni proposte nella prima parte della tesi e quelle che seguiranno: l'elaborazione di un piano d'indagine non può sottostare

alla logica che persegue l'adattamento della metodologia alla disponibilità economica, così da produrre pseudo-tecniche campionarie o modalità di raccolta dati che cancellano parte dei costi connessi alla qualità dell'impianto di ricerca. Il percorso seguito, al contrario, è stato quello che vede prima di tutto il discernimento delle soluzioni accettabili e, successivamente, la messa a punto di un'insieme di attività che vanno ad individuare uno spazio di soluzioni dove risultino minimi i costi metodologici, cioè i costi connessi agli obiettivi qualitativi (e/o scientifici) che è ragionevole poter raggiungere.

Una particolare attenzione è stata posta, e verrà posta anche nel proseguo, alle indagini CATI e all'uso degli *spare-panel*. Questo perché il CATI costituisce e ha costituito una modalità d'indagine così diffusa negli ultimi decenni da essere considerata quasi esclusiva e anche perché la costruzione dei piani di campionamento e la relativa conoscenza della popolazione da contattare (la lista) si è fondata fino ad oggi principalmente sugli elenchi telefonici. Il cambiamento tecnologico sta lentamente sbriciolando questa fondamentale "colonna portante" della ricerca campionaria (la conoscenza della "topografia" dell'universo di riferimento) tanto da far pensare che l'unica via per sostenere questo comparto produttivo sia quella di adattare "funzionalmente" anche la seconda "colonna portante", trasformare cioè la casualità nella scelta dei rispondenti in un "solido agglomerato di soggetti sempre pronti alla risposta", lo *spare-panel*. In questo modo, si pensa di superare i limiti connessi alla sovrapposizione degli "universi tecnologici", la resistenza alla risposta (rifiuti) e la lentezza con cui le persone si adattano alle nuove tecnologie. Del bisogno di condividere le tecnologie di comunicazione usate dalle persone (il telefono, internet, ecc.) e/o della necessità di predisporre strumenti adeguati alla raccolta dei dati (CATI, CAWI, ecc.), spesso se ne traggono alcune novità e le si eleva a rango di modernità. Nel contempo si oscura il pericolo di assistere ad una rovinosa caduta dell'"edificio scientifico" che è stato sviluppato sui concetti di casualità e pieno accesso

all'universo di riferimento. I vincoli dettati da ISSP, fortunatamente, portano a riflettere su questi temi e aprono nuovi orizzonti di confronto con la comunità scientifica internazionale circa le trasformazioni tecnologiche in atto. Ancora una volta il dibattito scientifico risulta fondamentale per non incorrere in possibili mistificazioni ed errori a cui le scienze sociali (così come quelle economiche) spesso incappano, sia dal punto di vista teorico sia rispetto agli strumenti. Il rischio è quello di giustificare la realtà piuttosto che governarla. È necessario far sempre collimare il percorso scientifico storico con un obiettivo: senza una rotta di navigazione, infatti, risulta inevitabile smarrirsi nel mare delle possibili direzioni.

Del percorso svolto a livello internazionale per alcuni tipi di attività di ricerca e di quello intrapreso per il progetto ISSP si tenterà di dare una descrizione più dettagliata nella terza e quarta parte della tesi.

3 – Geografie campionarie

3.1 Esperienze internazionali: dal CATI alle indagini postali

3.1.1 Il CATI: storia di un eldorado perduto

Il CATI, negli ultimi trent'anni, è stata la tecnica pressoché esclusiva nella ricerca sociale. L'indagine *face to face* e quella postale erano quasi del tutto scomparse, in quanto considerate:

- troppo onerose (per costi diretti e indotti)
- con tempi di sviluppo complessivi troppo elevati
- con tassi di adesione limitati (soprattutto la postale)
- affette da molteplici errori derivanti:
 - dal ruolo dell'intervistatore (soprattutto la *face to face*)
 - dall'elevato tasso di rifiuto e mancate risposte su singoli items (postale)
 - dalle problematiche di *routing* riscontrate nella compilazione del questionario (postale)
 - dalla tendenza all'autoselezione del rispondente all'interno della famiglia (postale).

La copertura dell'universo offerta dal telefono fisso è aumentata negli anni abbastanza repentinamente, seguendo di fatto un trend in linea con la crescente esigenza delle aziende commerciali di ottenere informazioni sulla popolazione. La polemica circa la parziale copertura del telefono fisso, si è quindi spenta abbastanza velocemente. In Italia, ad esempio, nel momento di massima espansione (attorno agli anni Novanta) è arrivata a superare in media il 95%. Attualmente, come si può vedere dai dati riportati nella Figura 3.1, la situazione è notevolmente cambiata.

Figura 3.1: Diffusione del telefono fisso in Italia
(estratto da una pubblicazione dell'autorità garante per le comunicazioni)

Tabella 2.19. Diffusione degli accessi alla rete fissa (dicembre 2011, in % delle famiglie)

Piemonte	66,3	Molise	58,5
Valle d'Aosta	59,1	Campania	70,9
Lombardia	71,0	Puglia	62,8
Trentino-Alto Adige	61,3	Basilicata	55,8
Veneto	60,5	Calabria	53,7
Friuli-Venezia Giulia	67,5	Sicilia	62,8
Liguria	70,9	Sardegna	59,3
Emilia-Romagna	67,3	ITALIA	67,4
Toscana	69,8	<i>Principali Comuni</i>	82,2
Umbria	62,8	<i>Nord Ovest</i>	69,6
Marche	66,5	<i>Nord Est</i>	64,1
Lazio	81,2	<i>Centro</i>	74,2
Abruzzo	61,7	<i>Sud e Isole</i>	63,4

Fonte: elaborazioni e stime su dati aziendali e Istat

Le percentuali riportate in Figura 3.1 sono da considerarsi, in ogni caso, delle sovrastime rispetto alla reale possibilità di contatto delle famiglie italiane. Infatti:

1. il processo di abbandono della telefonia fissa a favore di quella cellulare non si è ancora arrestato, sebbene sia notevolmente diminuito a seguito della diffusione dei collegamenti internet a banda larga su rete fissa
2. è incrementato notevolmente il numero di quanti oscurano il proprio numero telefonico negli elenchi pubblici (pagine bianche)
3. molti si sono iscritti al Registro delle Opposizioni⁷⁵ negando esplicitamente la disponibilità ad essere contattati telefonicamente per tutte le attività commerciali e/o di ricerca di mercato.

Su quest'ultimo aspetto, ovvero su cosa si debba intendere con il termine indagine, sondaggio di opinione o ricerca di mercato, è difficile

⁷⁵ «In conformità alle disposizioni del DPR n. 178/2010, pubblicato sulla Gazzetta Ufficiale n. 256 del 2 novembre 2010, gli Operatori che intendono contattare gli Abbonati presenti negli elenchi telefonici pubblici per attività commerciali, promozionali o per il compimento di ricerche di mercato tramite l'uso del telefono, sono tenuti a registrarsi al sistema gestito dalla Fondazione Ugo Bordoni - Gestore del servizio - e a comunicare la lista dei numeri che intendono contattare.» Tratto da <http://operatori.registrodelleopposizioni.it/>.

recuperare un giudizio condiviso e, soprattutto, in linea con la sensibilità della gente. Il telemarketing, teoricamente⁷⁶, dovrebbe far riferimento a questo Registro per filtrare i numeri telefonici con i quali operare; altrettanto dovrebbero fare coloro che si prestano a svolgere una ricerca di mercato (pur in assenza di un fine promozionale o commerciale diretto); giudizi contrastanti, al contrario, sembrano argomentarsi quando si parla di sondaggi d'opinione relativi, ad esempio, a temi politici. Tali sondaggi, ci si chiede, non servono molto spesso a definire al meglio il "prodotto" politico da offrire al cittadino? Un giudizio su di un candidato, un politico o un partito è forse diverso da quello su un *brand*, un modello d'auto o un servizio erogato a titolo oneroso e/o in ambito commerciale⁷⁷? Ad ogni modo, sta di fatto che «...sono oltre 1 milione i cittadini che hanno espresso il diritto di opposizione alle chiamate di telemarketing iscrivendosi al Registro Pubblico delle Opposizioni. A distanza di poco più di un anno dal lancio del servizio gratuito promosso dal Ministero...»⁷⁸. Cifra, che fa scendere di un ulteriore 4% (circa) i tassi medi riportati in Figura 3.1.

A forzare il declino del metodo CATI, sia in Italia che in quasi tutti gli altri paesi, non è quindi solo la scarsa copertura offerta dagli elenchi telefonici, ma anche una diminuita partecipazione alle indagini da parte

⁷⁶ Teoricamente perché vedremo poi come la norma venga di fatto violata.

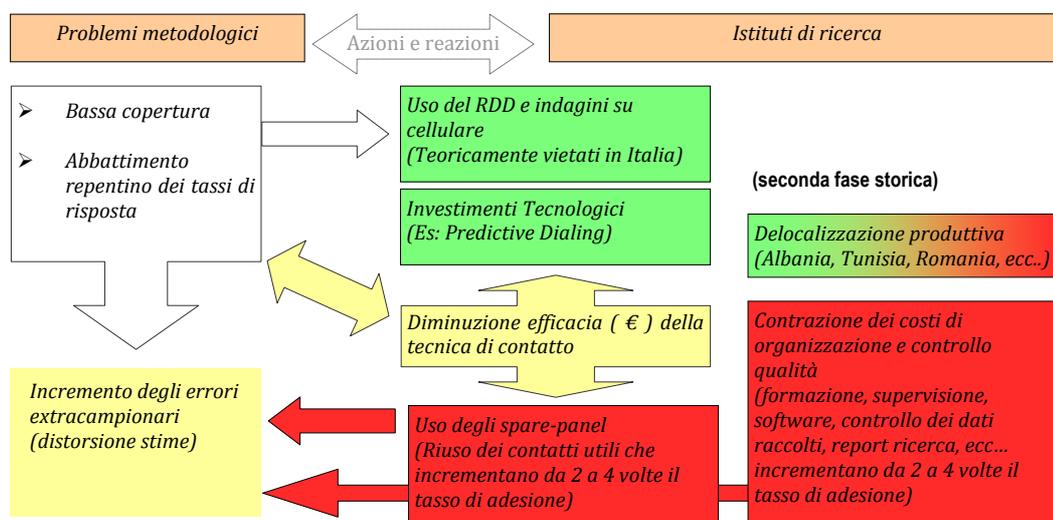
⁷⁷ Se si esula dalla stretta informazione relativa alle intenzioni di voto e dagli elementi socio-economici che descrivono al meglio il soggetto contattato (variabili dette di profilo), il nucleo principale della stragrande maggioranza dei questionari utilizzati nei sondaggi elettorali è orientato specificatamente al "marketing politico". La comprensione delle ragioni del voto e/o gli aspetti che più di altri intercettano i malumori dei cittadini, il classificare in ordine le necessità spicce/quotidiane della popolazione, sono colti al fine di approntare una campagna elettorale che enfatizzi le aspettative della gente e, come avviene nel marketing, siano utili a generare consenso attorno ad un candidato/partito. Lo strumento di ricerca, quindi, non è più orientato alla conoscenza delle ideologie presenti presso i vari gruppi sociali (credenze, opinioni, rappresentazioni, valori che orientano il comportamento di un determinato gruppo sociale), quanto al "consumo" di alcune parti costituenti l'ideologia stessa. Tutto questo rappresenta a mio avviso un elemento di equivalenza tra i sondaggi politici e le ricerche di mercato che esulano dalla conoscenza specifica del soggetto indagato, ovvero quelle che concentrano la loro attenzione sugli elementi "ideologici" del consumo di un bene. Da questo punto di vista, la distinzione sull'argomento d'indagine non può essere considerato lo spartiacque tra "ricerca di mercato" e "sondaggio d'opinione". Eppure, una ricerca di mercato, seppur anonima, espone la società di ricerca alla possibile contestazione da parte dell'autorità garante.

⁷⁸ Tratto dal <http://www.registrodelleopposizioni.it/it/news> (Registro delle Opposizioni, News del 18 Aprile 2012).

della popolazione; causa l'assillante frequenza di chiamate dettate dallo sviluppo del telemarketing e, in fondo, l'abuso stesso della tecnica CATI⁷⁹. Da oltre dieci anni, il proverbiale rapporto tra costi e benefici del CATI si è talmente ridotto da far bandire la tecnica stessa da molte indagini sociali a carattere scientifico. Rimane comunque un metodo molto veloce e, proprio per questo, il suo utilizzo rappresenta ancora oggi la pratica più diffusa per le indagini commerciali, politiche e comunque spesso inserito in indagini più articolate e svolte con tecniche miste.

Il cambiamento di scenario, che con tempi e intensità diverse si è venuto a creare in molti paesi, porta con sé problematiche e reazioni sia da parte del mondo scientifico, che da parte del comparto produttivo. Per comprendere al meglio quali siano le ansie del primi e le reazioni del secondo, forse è utile affidarsi allo schema di seguito esposto.

Figura 3.2: Schema di riferimento sulle trasformazioni del CATI



Quale risposta al problema della bassa copertura (e, durante la fase di espansione della telefonia fissa, di non avere a disposizione una lista completa ed aggiornata delle utenze) si è da sempre adottata la tecnica

⁷⁹ Il prof. Bernardi (2005) ha parlato di "molestia statistica" riferendosi all'abbondanza di azioni d'indagine presso la popolazione

RDD (*Random Digit Dialing*)⁸⁰ e, successivamente, si è iniziato a contattare gli intervistati al cellulare⁸¹. Per contrastare l'abbassamento dei tassi di risposta, ossia la diminuita efficacia produttiva, si è massicciamente investito sulla tecnologia che vede l'integrazione del computer con il telefono (CTI - Computer Telephone Integration - con funzionalità di *predictive dialing*⁸²). Queste risposte tecniche e di metodo (RDD, indagini su cellulare e *predictive dialing*) rispondono in parte alla necessità di abbattere i potenziali errori extracampionari, ossia rappresentano prassi metodologicamente adeguate, ma non sono strettamente in linea con le normative in materia di protezione dei dati personali (d.lg. n. 196 del

⁸⁰ Il *Random Digit Dialing* è una tecnica nata per la generazione di numeri telefonici da associare ai diversi territori (prefissi). Tradizionalmente, lo sviluppo di una rete telefonica (all'interno di una nazione) avviene suddividendo il territorio in svariate aree (identificate dai prefissi numerici: XXX) e installando, all'interno di ognuna, diverse centrali telefoniche (identificate dai suffissi numerici: YYY). Infine, le risorse di centrale (linee) vengono allacciate alle diverse utenze finali (identificate da un ulteriore numero ZZZZ). Il numero da comporre per chiamare l'utente è quindi il seguente: XXX-YYY-ZZZZ, ossia prefisso-suffisso di centrale-numero utenza. In un periodo di pieno sviluppo della telefonia fissa, così come quando non esiste una lista telefonica pubblica, la conoscenza dei prefissi e dei suffissi telefonici (XXX,YYY) permette la generazione di numeri telefonici semplicemente creando una stringa numerica casuale di lunghezza nota (la parte ZZZZ del numero). Per la telefonia cellulare in Italia, fino al momento in cui furono emanate le delibere della Commissione per le Infrastrutture e le Reti (n. 12/01/CIR; n. 19/01/CIR) che permettono la portabilità del proprio numero, il prefisso identificava l'operatore telefonico. La relazione tra territorio e utente è venuta quindi a cadere con lo sviluppo della telefonia cellulare e con l'introduzione della tecnologia VoIP anche se permane a tutt'oggi abbondantemente nella telefonia fissa.

⁸¹ In Italia, la presenza di elenchi pubblici e unitari per la telefonia su rete fissa ha reso pressoché inutilizzata la tecnica RDD applicata al CATI. Similmente, nei paesi in cui chi riceveva la telefonata sul mobile pagava parte del costo (come in uso anni fa negli Stati Uniti), le indagini su cellulare si sono scarsamente diffuse. In ogni caso, entrambe queste "risposte" di metodo, in Italia sono oggi teoricamente vietate (si veda a tal proposito la Newsletter n. 332 del 10/12/2009 in www.garanteprivacy.it).

⁸² Sulla base di numerosi parametri (collegati alla durata del contatto, al numero di operatori al lavoro, il loro stato di disponibilità a un successivo contatto, alla storia recente e pregressa delle chiamate relative a quel progetto, ecc.) il sistema telefonico prevede un numero superiore di chiamate telefoniche da lanciare rispetto al numero di operatori disponibili in quel momento. In questo modo gli operatori sono "serviti" da una telefonata (contatto reale) in maniera più veloce e non rimangono a lungo in attesa che qualcuno risponda al telefono. Il sistema, quindi, anticipa le chiamate rispetto al momento in cui l'operatore sarà effettivamente disponibile, così come ne "abbatte" alcune se queste sono state generate in esubero, cioè se il numero di quanti rispondono al telefono risulta superiore rispetto agli operatori disponibili in quel momento. Queste sono le chiamate "senza risposta" che a volte riceviamo a casa. In molti paesi queste tecnologie sono regolamentate al fine di non creare eccessivo disturbo alla popolazione, in modo tale che le telefonate in uscita siano in larghissima parte (95% generalmente) "coperte" da un operatore disponibile piuttosto che inserite in "attesa" o "abbattute". In Italia solo recentemente il Garante per la protezione dei dati personali si è pronunciato su tale argomento (si veda chiamate "mute" su Registro dei provvedimenti n. 474 del 6 dicembre 2011 e n. 83 del 20 febbraio 2014).

30/06/2003). Diverso, invece, è l'uso degli *spare-panel* e la contrazione dei costi transazionali di produzione dell'informazione e di controllo della qualità dei dati raccolti. Sul loro inappropriato effetto, abbiamo già argomentato nella prima parte della tesi. Della delocalizzazione produttiva, spinta in modo oramai travolgente grazie all'utilizzo della tecnologia VoIP, lasciamo al lettore ogni giudizio. Rimane chiaro, tuttavia, che sia la delocalizzazione produttiva sia la diminuita opportunità d'agire con la tecnica CATI-RDD⁸³, produce una perdita di competenze e un disinvestimento in termini di studio e sviluppo tecnologico. Inoltre, l'azione congiunta dell'utilizzo degli *spare-panel* telefonici e l'abbattimento anche del 70% dei costi attraverso la delocalizzazione, rende pressoché incolmabile il gap economico a carico degli istituti che in Italia volessero produrre ricerca metodologicamente corretta (Brusati 2001). In tutto questo, il paradosso eclatante è il fatto che l'aver spostato all'estero le attività di rilevazione consente ai più di continuare a eludere il divieto di fare indagini di mercato (propriamente dette) e azioni promozionali (telemarketing) senza prima passare per il Registro delle Opposizioni. Si evita così il pagamento del servizio di filtro sui numeri telefonici che si

⁸³ Nel provvedimento di divieto, l'Autorità garante ha spiegato che anche il numero casualmente composto deve considerarsi quale "dato personale", in quanto ricollegabile, anche indirettamente, a una persona identificata o identificabile. Si afferma, relativamente al caso di un operatore che usava la tecnica del RDD per la creazione dei numeri telefonici: "...considerato che lo svolgimento, da parte della società, di chiamate indesiderate verso numeri di destinatari, ottenuti con digitazione casuale e senza previo consenso, è da ricondurre all'ambito applicativo in questione (leggi: "violazione circa le norme di tutela dei dati personali"), altrimenti verificandosi una elusione degli adempimenti in materia di informativa e consenso di cui agli artt. 13, 23 e 130 del Codice" (Provvedimento di divieto del Garante della Privacy del 03/10/2009).

usano⁸⁴ e si continua a sollecitare oltremodo la popolazione con tentativi di vendita di prodotti.

3.1.2 Verso l'Address Based Sampling (ABS)

Di fronte all'inesistenza di una tecnica altrettanto adeguata alle necessità della ricerca scientifica come lo è stato il CATI, nell'ultimo decennio molti istituti di ricerca esteri (prevalentemente statunitensi) hanno optato per:

- implementare campionamenti non più basati su elenchi telefonici, ma su altre fonti anagrafiche,
- adottare tecniche ibride (si sondano via CATI le unità campionarie "coperte" dalla telefonia fissa; le restanti sono raggiunte a casa dal rilevatore o contattate via posta) in modo tale da raggiungere la massima copertura campionaria possibile,

ovvero, hanno scelto di ritornare alle più economiche indagini postali, sviluppando piani di campionamento basati sugli indirizzi (*Address Based Sampling*). A differenza di un tempo, infatti, l'informatica è ormai entrata

⁸⁴ Se si stipula un accordo commerciale di intermediazione tra il committente e la società di telemarketing straniera, i committenti si liberano giuridicamente dell'onere legato a come i propri partner commerciali procurano i clienti finali (consumatori) e, di conseguenza, questo genera un alibi rispetto al doversi sottoporre alla normativa riguardante il Registro delle Opposizioni. Per le indagini di mercato frapponendo, tra il soggetto contattato e il committente dell'indagine, un'azienda estera che effettua la raccolta dei dati, il Registro delle Opposizioni non è quasi mai utilizzato. Tale Registro, pur nella sua necessità d'essere, ha di fatto reso ancora più confusa l'idea che le persone hanno della possibilità d'essere contattate telefonicamente, ossia la distinzione tra l'essere oggetto di un'attività di telemarketing, ricerca di mercato o sondaggio d'opinione. Tant'è, che anche nel caso si facciano sondaggi di opinione, ci si trova spesso di fronte a cittadini che paventano azioni legali proprio in ragione della loro presunta tutela fornita dall'essere iscritti al Registro delle Opposizioni. Lo spostamento all'estero delle attività di raccolta dei dati e la reale difficoltà da parte del Garante di perseguire tali azioni, ha creato ulteriori difficoltà alle aziende italiane del settore. È bene ricordare che in altri paesi le normative e le situazioni di mercato sono differenti. Ad esempio, è regolamentato l'uso del *predictive dialing* (in Italia solo recentemente, a fine 2013, il Garante si è pronunciato in merito). Ad esempio, negli Stati Uniti è permesso l'uso del RDD. È distinto in modo più appropriato ciò che fa riferimento alla ricerca (da intendersi anche quale strumento al servizio dello sviluppo commerciale delle aziende) da quello che sono le azioni di vendita o di promozione diretta dei prodotti (pubblicità), ecc. In Italia, purtroppo, si è trasformata una lecita richiesta da parte dei cittadini in un ulteriore adempimento burocratico oneroso per le aziende di sondaggi. Si sarebbe potuto predisporre un servizio telematico gratuito per verificare le utenze che esplicitamente hanno espresso la volontà di non essere contattate in alcun caso. La confusione tra quello che è possibile svolgere in termini di contatti telefonici e quello che non sarebbe possibile (ma che di fatto avviene), rimane elevata. Un motivo in più, si mormora nel settore, per trasferire le proprie attività all'estero.

prepotentemente nel campo della logistica e, oggi, in molte nazioni sono presenti dei buoni archivi contenenti gli indirizzi delle abitazioni e delle attività produttive, gestiti da software che permettono il calcolo ottimizzato dei percorsi da seguire per le consegne (DSF - *Delivery Sequence File*). Queste informazioni, spesso in possesso delle società che erogano il servizio postale, hanno rappresentato una vera e propria innovazione di processo (produttivo) e sono state colte dai ricercatori come una valida opportunità per la generazione di campionamenti areali basati sulla scelta casuale di soggetti in partizioni territoriali specifiche.

Al fine di ottimizzare la copertura campionaria, molti istituti di ricerca hanno messo a punto svariati metodi di *fieldwork*, ossia tecniche di perlustrazione delle aree interessate dall'indagine, tali da controllare e reperire la lista completa delle abitazioni presenti sul posto (Iannacchione et al., 2012; Ridenhour et al. 2008). Da qui, poi, la scelta casuale delle famiglie e soggetti da contattare⁸⁵. Uno degli elementi interessanti di queste esperienze è il fatto che alcuni studi comparativi svolti negli USA hanno stabilito che esiste un'equivalenza, se non addirittura un vantaggio relativamente ai costi totali dell'indagine e ai tassi di adesione (attorno al 26-27%), tra le indagini basate sugli indirizzi (ABS-DSF) e quelle svolte con tecnica CATI e RDD (Battaglia et al., 2008). Costi e tassi di adesione sono due entità analitiche che molto hanno a che fare con i territori sui quali si indaga; parlando di costo metodologico e riportando un esempio numerico, tuttavia, abbiamo visto che anche per l'Italia il bilancio economico-metodologico non è poi dissimile dai risultati a cui arrivano gli studi citati.

3.1.2.1 Panel e campioni sperimentali istantanei: diversi strumenti per diversi obiettivi

Prima di approfondire gli aspetti positivi e negativi delle ricerche che si possono implementare grazie alla conoscenza degli indirizzi, è forse il

⁸⁵ Ricordiamo che in alcune nazioni, come gli USA, non esistono o non sono facilmente reperibili anagrafi della popolazione residente in un determinato territorio amministrativo. La necessità di una enumerazione delle abitazioni attraverso il presidio fisico del territorio diventa quindi quasi inevitabile.

caso di sottolineare ancora una volta come il percorso razionale attivato dai vari istituti di ricerca scientifica sembra aver scartato l'ipotesi di costruire dei panel da contattare tramite il CAWI. La necessità di coprire l'intero universo campionario, di focalizzare gli studi su determinati territori d'interesse, di controllare nel tempo la qualità del proprio panel⁸⁶, ha circoscritto un percorso metodologico che della modernità tecnologica ha colto essenzialmente la sola "geografia informatizzata", ossia la capacità computazionale e la possibilità di immagazzinare una notevole quantità di dati. In altri termini, si è preferito ritornare a pratiche di ricerca che possono sembrare vetuste (come l'indagine postale), ma sulle quali la quantità di lavoro a posteriori per le verifiche di qualità è più contenuta rispetto a quella necessaria per l'implementazione di solidi panel. Ovviamente, gli strumenti adottati (in questo caso i panel piuttosto che i campioni sperimentali istantanei) non possono essere scambiati a piacimento e in modo indipendente rispetto agli scopi dell'indagine. Ottenere risultati sui cambiamenti nel tempo degli oggetti indagati (longitudinali su panel) è un obiettivo ben diverso dal fornire un'istantanea del panorama sociale che si ha di fronte in quel momento (campioni sperimentali istantanei). Nulla vieta che i due metodi possano coesistere all'interno dello stesso esperimento. Anzi, questa prospettiva rappresenterebbe un accrescimento di conoscenza.

La convenienza economica nell'adottare un panel quale strumento per l'analisi del contingente è da calcolarsi in funzione del rapporto tra il risparmio indotto dalla disponibilità alla risposta dei soggetti appartenenti al panel e i costi necessari per implementare i campioni di "controllo e rabbocco" del panel stesso. Il panel parte dall'idea di tenere fermo una

⁸⁶ Molto spesso un panel è costruito a partire da una tecnica probabilistica di scelta delle unità che ne faranno parte. Tuttavia, se si vuole mantenere la rappresentatività nel tempo del campione-panel rispetto alla popolazione generale, sono necessarie diverse azioni d'indagine sperimentale, nonché l'eventuale introduzione nel panel stesso di ulteriori soggetti provenienti da campioni rappresentativi. In sintesi, per mantenere rappresentativo e sperimentale nel tempo un panel, bisogna attivare tutte le pratiche tipiche delle indagini longitudinali con rotazione (Kasprzyk et al., 1989). In questo modo, però, diminuisce sensibilmente il vantaggio economico.

porzione dell'iniziale "spazio" campionario, mentre tutto il resto (l'universo/l'urna) nel tempo si muove, definendo inevitabili nuovi scenari. Un'idea decisamente brillante, ma soggetta a:

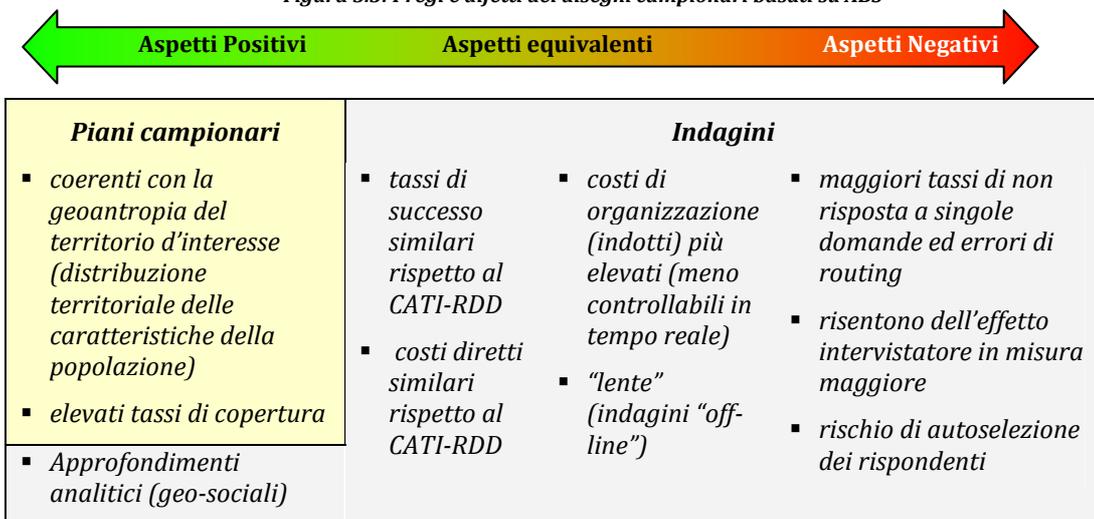
1. perpetrare nel tempo gli eventuali errori di copertura iniziali
2. assommare ulteriori errori di copertura in base alla stabilità numerica del panel nel tempo (disomogeneità delle "cadute" in funzione di alcune caratteristiche della popolazione-panel)
3. accumulare ulteriori distorsioni tanto più risulta dinamico il panorama circostante (dinamica esogena al panel e riferibile all'universo di riferimento) e in base a quanto l'appartenere ad un panel modifichi il modo di rispondere dei soggetti indagati (dinamica endogena al panel).

Sia la costruzione del panel sia la sua "tenuta" nel tempo, necessita quindi di una cura e di una attenzione particolare. Inoltre, esso non è adatto per ogni spazio di campionamento, ad esempio nel caso in cui si debba (o sia più conveniente) eseguire un'indagine su particolari sub-territori.

3.1.3 Pregi e difetti dei disegni d'indagine basati sugli indirizzi

Passare da indagini basate su collegamenti telefonici a quelle su indirizzi porta con sé pregi e difetti.

Figura 3.3: Pregi e difetti dei disegni campionari basati su ABS



Gli aspetti decisamente positivi connessi ad una opportuna conoscenza della lista degli indirizzi riguardano prevalentemente l'ambito del campionamento. Si è già citata la possibilità di ottenere tassi di copertura molto elevati, almeno tanto quanto lo erano gli elenchi telefonici di un tempo. Il vero elemento distintivo dell'indirizzo sta però nella possibilità di un suo posizionamento geografico: la mappatura sul territorio dell'insieme degli indirizzi consente di ricostruire esattamente la struttura antropica e geo-sociale dell'universo d'indagine. La densità di popolazione presente sul territorio può essere ricostruita adeguatamente proprio a partire dagli indirizzi delle abitazioni; così come possono essere "mappate" tutte le caratteristiche conosciute a priori o dopo la rilevazione dei dati.

Il numero di telefono di rete fissa, contrariamente all'indirizzo, è solo un riflesso della distribuzione della popolazione nel territorio. Può essere (nel tempo) scambiato da utente ad utente e la centrale telefonica può gestire migliaia e migliaia di utenze, coprendo così territori generalmente più ampi di un quartiere, trasversali a determinate suddivisioni amministrative (province e comuni) e sparsi in aree più grandi delle sezioni censuarie (o similari ripartizioni in altri stati). L'indirizzo, al contrario, consente di definire pressoché a piacimento l'ampiezza (la "granularità") della ripartizione territoriale desiderata. Dato il suo scopo connesso alla reperibilità di una famiglia o di un soggetto specifico, potenzialmente fino ad oltre il singolo edificio. Il numero telefonico, da questo punto di vista, rappresenta solo uno dei tanti attributi collegati/collegabili all'indirizzo-utente.

Per far comprendere meglio gli aspetti positivi nell'utilizzo degli indirizzi, forse è il caso di citare le linee di sviluppo di un'indagine che lo IUAV di Venezia ha commissionato nel (2007) alla società Marker Srl di Mestre Venezia, dove si sono reclutate telefonicamente delle persone disposte a rispondere, presso una sede adibita allo scopo, ad un questionario informatizzato che riportava diversi elementi grafici. Oggetto dell'indagine era l'atteggiamento e le opinioni dei cittadini circa le modalità

e le opportunità d'intervento nelle ex aree industriali inquinate di Porto Marghera. Il campione suddivise i rispondenti in gruppi definiti in base alla distanza tra residenza e aree in oggetto. In questo caso, potremmo dire, è stato il campione che ha preso forma dal territorio. Il reperimento telefonico dei partecipanti, infatti, è avvenuto a fronte della conoscenza dell'indirizzo sul quale poggiava l'utenza. Questo ha permesso di avere una numerosità adeguata per tre fasce di distanza, considerate a priori dei possibili "contenitori" di differenziate "opinioni medie". La conoscenza degli indirizzi, in questi casi, permette di ampliare le possibilità interpretative dei fenomeni rilevati⁸⁷.

Gli indirizzi, in generale, non sono concepiti per essere facilmente riversati sul piano delle ripartizioni amministrative (sezioni di censimento, quartieri o altro)⁸⁸. Invece, per facilitare la consultazione, le liste telefoniche su rete fissa sono state concepite suddivise per porzioni territoriali ampie, ma ben definite (in Italia i comuni). La costruzione della geografia del territorio attraverso gli indirizzi diventa quindi un'esigenza primaria per poter effettuare indagini postali o dirette con campioni areali. Il processo di posizionamento nella mappa territoriale degli indirizzi è detto geocodifica.

3.1.4 ABS e Geocodifica

3.1.4.1 La geocodifica per la selezione della lista campionaria

Spesso le indagini coinvolgono non tanto l'intera popolazione del territorio nazionale, quanto i residenti di una partizione più limitata (un quartiere, un comune, una provincia, ecc.). Similmente, se si ritiene opportuno definire un campione stratificato o a stadi rispetto alla popolazione presente in diverse ripartizioni amministrative, si sottende

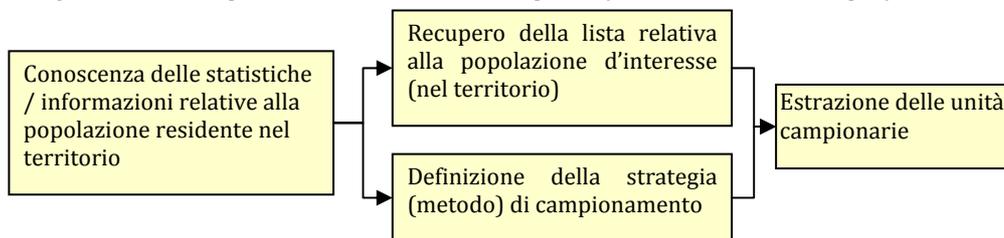
⁸⁷ In ambito commerciale tali analisi sono identificate come studi di geomarketing.

⁸⁸ A volte, nemmeno il loro corretto posizionamento all'interno di un comune non è così scontato, data la presenza di numerose frazioni, la possibile mancanza di alcune indicazioni di dettaglio circa le aree CAP e altro.

l'esistenza di ragionevoli ipotesi che i territori stessi abbiano un ruolo attivo nella comprensione delle diversità (variabilità) rilevabili⁸⁹.

In generale, qualsiasi sia la lista di riferimento (elenco telefonico, anagrafe ...), il processo di creazione di un campione prende spunto, se disponibili, da informazioni statistiche (spesso censuarie) sull'universo di riferimento. Dopo aver recuperato la lista della popolazione, si definisce (molte volte congiuntamente) la tecnica di campionamento. Solo alla fine, applicando una selezione casuale dalla lista, si arriva a definire l'elenco delle unità da contattare.

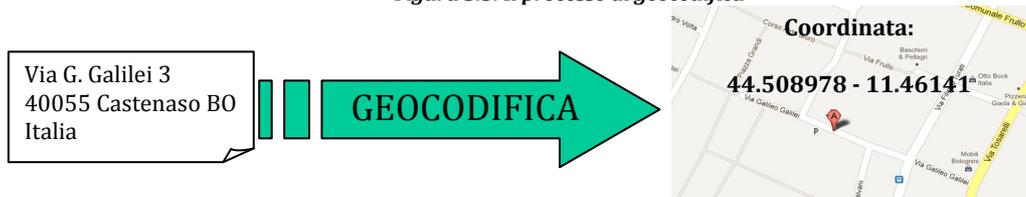
Figura 3.4: Classico percorso di creazione di un campione riferibile ad un territorio specifico



Qualora la lista sia un indirizzario, per costruire un campione relativo ad un determinato territorio e analizzare i dati della popolazione sondata a livello geografico, è necessario applicare la geocodifica agli indirizzi. Dal punto di vista tecnico si tratta di trasformare l'indirizzo in una coordinata; dal punto di vista sostanziale, verificare che l'unità sia pertinente all'universo di riferimento. La sovrapposizione tra il *layer* definito da punti (le coordinate degli indirizzi geocodificati) e il *layer* areale definito dai poligoni che rappresentano il territorio d'interesse, permette di isolare gli indirizzi che vi appartengono. Questa operazione, che sfrutta il processo di geocodifica, è realizzabile tramite gli strumenti inseriti nei software GIS (*Geographic Information System*).

⁸⁹ A volte, più semplicemente, si tratta di applicare un principio di cautela, in assenza di informazioni specifiche circa l'effetto della struttura antropica dei territori indagati sul fenomeno di interesse. Ad esempio, una provincia viene suddivisa tra comuni scarsamente popolati (piccoli), medie comunità e grandi aggregati urbani. Ancora, comuni di pianura, collinari e/o di montagna; ecc.

Figura 3.5: Il processo di geocodifica



3.1.4.2 Qualità dei DSF e qualità del processo di geocodifica

L'operazione di geocodifica non è avulsa da errori. Il dibattito scientifico internazionale ha spinto affinché siano valutati gli errori di sovra/sotto copertura degli universi dovuti all'utilizzo dei DSF e della geocodifica quale strumento di determinazione della lista (universo) su cui basare l'estrazione del campione (Eckman 2012; Eckman e English, 2012a, 2012b; Iannacchione et al., 2010a, 2012b). Negli USA, a seconda del fornitore del *file*, si sono raggiunti vari risultati. In linea generale, il tasso di copertura dell'universo offerto dai DSF è variabile da territorio a territorio. Le aree rurali sono quelle che presentano minore copertura. Per le aree urbane la copertura è pressoché totale⁹⁰, se non addirittura ridondante. Mediamente, la copertura è stimata attorno al 92,3%. Alla luce dei risultati raggiunti, molti istituti di ricerca ritengono opportuno effettuare dei sopralluoghi solo nelle aree caratterizzate da basse coperture e solo quando il recapito del questionario preveda la consegna dello stesso da parte di rilevatori. Negli altri casi (come l'invio di questionario postale entro micro-aree), il file DSF può essere considerato soddisfacente. Sempre negli USA, nell'83,3% dei casi il processo di geocodifica è stimato "molto preciso" (a livello di blocco/tratto censuario), nel 16,5% la precisione è da considerarsi "bassa" (si individua l'aggregato territoriale di livello superiore)⁹¹; solo nello 0,2% gli indirizzi non trovano una loro geocodifica. Se ciò rappresenta la realtà, significa che in oltre l'80% dei casi l'indirizzo è correttamente posizionato nella sezione di censimento, l'aggregato territoriale più piccolo sul quale è possibile disporre di informazioni censuarie sulla popolazione. Una "granularità" informativa che permette

⁹⁰ Il tasso di copertura è al 60% circa in Alaska; 97% nell'Ohio e Rhode Island.

⁹¹ Da non dimenticare che negli USA è molto diffuso, soprattutto nelle aree rurali, l'utilizzo delle "box-office", ossia il ricevimento della propria corrispondenza presso caselle postali.

decisi incrementi sulle opportunità di campionamento e la definizione di piani di analisi (come quelli relativi alla pianificazione sociale/urbana) alquanto sofisticati. È questa l'opportunità offerta dall'analisi geo-sociale basata sugli indirizzi.

L'utilizzo dei DSF congiuntamente alla geocodifica permette la riduzione sostanziale delle attività sul campo che, precedentemente, hanno costituito la regola (e incrementato i costi) di ogni indagine postale o *face to face* su territori selezionati. Grazie alla geocodifica, la predisposizione di una lista pertinente per un determinato universo/territorio è resa più efficiente. L'abbattimento dei costi così ottenuto, permette alle indagini postali e dirette di assumere nuove caratteristiche e di diventare economicamente più convenienti rispetto al recente passato. Rimangono ovviamente diversi punti critici, ma siamo di fronte a una soluzione metodologicamente corretta rispetto alla necessità di salvaguardare la casualità nella scelta del rispondente e alla potenziale completa copertura dell'universo di riferimento. Lo strumento di geocodifica applicato ai DSF rappresenta una vera e propria innovazione di processo: permette il superamento della tecnica CATI-RDD che riscontra tassi di risposta sempre più limitati e che sta accumulando problematiche di copertura campionaria.

3.2 Replicabilità delle indagini ABS in Italia

3.2.1 Assenza dei file DSF e le fonti anagrafiche in Italia

In Italia non esistono file simili al DSF e, da quel che risulta, la geocodifica delle anagrafi comunali non è mai stata tentata. I fornitori di elenchi telefonici, al contrario, hanno geocodificato gli indirizzi associati alle (oramai scarse) utenze telefoniche di rete fissa, ma non è nota la qualità del risultato ottenuto⁹². L'assenza dei DSF in Italia, così come la lentezza delle indagini dirette, ha certamente inciso sulla deriva sperimentale delle pratiche di ricerca. In tal senso, non è nemmeno del tutto errato affermare che la vera novità rappresentata per oltre trent'anni dalla tecnica CATI non è stata tanto la tecnologia ad essa connessa (il telefono ed il computer), quanto il progressivo accumulo d'informazione rappresentato dagli elenchi telefonici. Infatti, che si tentasse un'indagine CATI o un'indagine postale o *face to face*, l'elenco telefonico (specie informatizzato) ha sempre rappresentato un ausilio prezioso.

Data l'assenza di file simili ai DFS, nell'impianto dell'indagine ISSP Italia non si è inizialmente fatta particolare attenzione alle esperienze internazionali che portavano a valutare l'efficienza delle indagini postali. In Italia, per l'appunto, le indagini basate sugli indirizzi sono costruite a partire dalle anagrafi comunali o su liste elettorali⁹³, informazioni ritenute particolarmente onerose. Le anagrafi comunali, particolarmente complete e precise, sono pubbliche, ma dichiarate "accessibili e rilasciabili" dall'Amministrazione Comunale solo per "pubblica utilità" verso altre Amministrazioni pubbliche (capo VI art. 34 del D.P.R. n. 223/1989) e fornite dall'Ufficiale dell'anagrafe solo ed esclusivamente se le informazioni sono rese anonime e in forma aggregata per fini statistici e di ricerca.

⁹² È mia personale opinione/esperienza che la situazione in Italia non differisca di molto da quella presente negli USA.

⁹³ Esiste un prodotto postale che consente la consegna "non indirizzata" a tutte le famiglie di una zona (ad esempio un'area CAP). Si tratta di attività di consegna adatta a grandi volumi e generalmente limitata ad usi pubblicitari. L'impossibilità del controllo dell'attività sul campo rende lo strumento inadatto a scopi di ricerca.

Viceversa, le liste elettorali, contenenti l'elenco della popolazione maggiorenne residente nel comune e in possesso dei diritti politici⁹⁴, sono accessibili, ma la loro consultazione è spesso problematica per difficoltà pratiche e procedurali e, molto più spesso, per inaccettabilità delle finalità per le quali se ne chiede la visione⁹⁵. Anche quando possibile, quindi, l'accesso a queste fonti genera un costo alquanto elevato sul *budget* complessivo dell'attività di ricerca, nonché ritardi a volte insostenibili (cfr. par. 2.3.2).

Il superamento dell'utilizzo dell'elenco telefonico da un lato, la difficoltà a reperire altre fonti anagrafiche dall'altro e, di conseguenza, la difficoltà a riproporre in Italia le esperienze internazionali, ha trovato sintesi nell'approfondimento delle potenzialità connesse alla geocodifica (funzione di *reverse geocoding*).

3.2.2 Il Reverse Geocoding e la proposta di campionamento per l'Italia

Nel processo di geocodifica, come già ribadito, l'attenzione dei ricercatori si è focalizzata nel comprendere quali errori di sotto-sovracampionamento si possono commettere utilizzando i file DSF. Nessuno, fino a poco tempo fa, si era soffermato a valutare le opportunità che offre la funzione di *reverse geocoding*, cioè la possibilità di farsi restituire un indirizzo a partire da un punto definito sul territorio. Questo perché:

1. il processo di geocodifica non è una funzione "biunivoca": il *reverse geocoding* restituisce al più un tratto di strada (compreso in un intervallo di civici) e quasi mai un singolo indirizzo
2. la generazione di punti distribuiti omogeneamente e in modo casuale su un territorio, vale a dire la restituzione di un set di indirizzi a

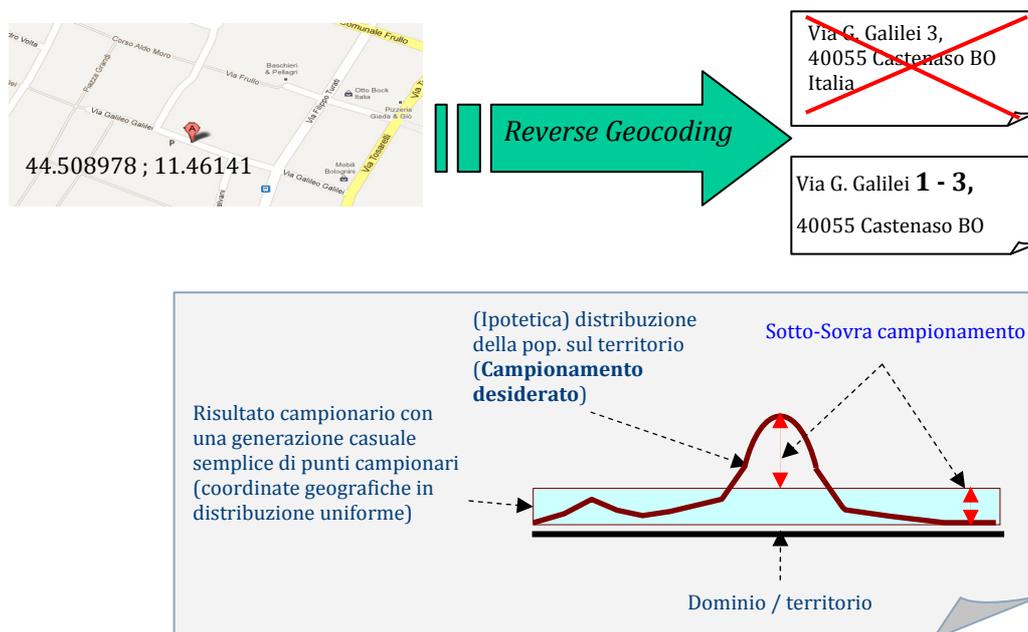
⁹⁴ L'indagine ISSP è rivolta alla popolazione residente in età superiore ai 16 anni. Le liste elettorali sarebbero quindi deficitarie di una parte, seppur limitata, dell'universo di riferimento.

⁹⁵ "Le liste elettorali possono essere rilasciate in copia per finalità di applicazione della disciplina in materia di elettorato attivo e passivo, di studio, di ricerca statistica, scientifica o storica, di analisi di carattere socio-assistenziale o per scopi di interesse collettivo o diffuso" (art. 177, comma 5 del d.lg. 30 giugno 2003, n. 196 "Codice in materia di protezione dei dati personali").

partire da coordinate geografiche, porterebbe ad un sovracampionamento delle aree scarsamente popolate (e viceversa). Il processo di *reverse geocoding*, in altre parole, non è in grado di fornire alcuna indicazione sullo sviluppo verticale dell'aggregato urbano

- il tratto stradale restituito, comprensivo eventualmente dell'intervallo di numeri civici che lo identifica, potrebbe essere esterno all'area d'interesse.

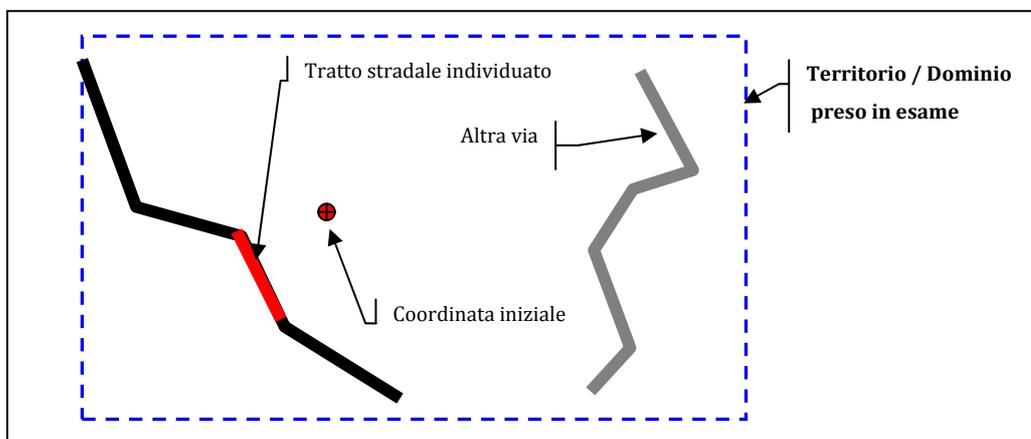
Figura 3.6: Il processo di reverse geocoding: tipo di dati restituiti e problema del sovra-sotto campionamento



Inserendo una coordinata all'interno di un territorio, viene restituita la strada più vicina, quella a distanza minima dal punto. Sebbene la coordinata iniziale appartenga ad un'area definita, non è detto che ad essa appartenga anche il tratto di strada restituito dalla funzione di *reverse geocoding*. Tale tratto stradale (generalmente definito da due incroci o entro due punti particolari quali un edificio specifico ed un incrocio, due edifici specifici, o altro) è a sua volta qualificato da molteplici altri attributi (denominazione completa, tipo di strada, senso di marcia, tempi di

percorrenza, ecc.) che purtroppo a volte sono assenti⁹⁶. Una volta inserita una coordinata geografica, il tratto stradale restituito potrebbe non essere accompagnato neppure dalla denominazione completa: ciò dipende dalla qualità dei *database* geografici sottostanti. Quando invece si ottiene un intervallo di numeri civici, come nella maggior parte dei casi, la correttezza del reale posizionamento di un edificio lungo il tratto stradale è funzione dell'omogeneità con cui i civici stessi si dispiegano lungo la via. Tanto più gli edifici sono ben distribuiti nel tratto stradale, tanto più preciso è il punto in cui la funzione di *reverse geocoding* "posiziona" l'edificio stesso.

Figura 3.7: Processo di reverse geocoding: determinazione del tratto stradale attraverso una coordinata



Per le problematiche appena evidenziate e avendo la possibilità di disporre del file DSF, negli USA la funzione di *reverse geocoding* fino a poco tempo fa non è mai stata presa in considerazione come soluzione facile e praticabile nei casi di indagine postale. Tale opportunità, consente invece la convergenza dell'esperienza di ricerca italiana con quelle internazionali. Vediamo in che modo.

3.2.2.1 L'individuazione del civico dove recapitare il questionario

Per gli USA è stato dimostrato come in circa oltre l'80% dei casi il processo di geocodifica offre un risultato eccellente posizionando un

⁹⁶ Molte sono le geometrie e i dati generalmente presenti e restituiti dalla funzione di *reverse geocoding*. La completezza e l'aggiornamento degli elementi di qualificazione, però, non sono sempre assicurati. La qualità dei *database* geografici, per meglio dire, non è omogenea in tutte le parti del mondo e nemmeno all'interno di una stessa nazione. Non esiste alcun organo nazionale che governi questo tipo di informazioni.

indirizzo esattamente nel tratto/blocco censuario pertinente. Ciò significa che, fornito un indirizzo (via e civico), la funzione di geocodifica lo interpreta prima di tutto come appartenente ad un preciso tratto stradale. Questo, d'altra parte, è ciò che succede quando si chiede ad un navigatore satellitare di definire il miglior percorso verso una determinata meta. Facciamo ora il ragionamento inverso. Dato un tratto stradale (anagrafica del *database* geografico) è possibile isolare la porzione di territorio per la quale quel determinato tratto stradale rappresenta "la via più vicina". In altri termini, per ogni punto geografico appartenente a quell'area la funzione di *reverse geocoding* fornirà come risultato il tratto stradale inizialmente isolato. Per oltre l'80% dei casi, quindi, al netto dei sovra-sotto campionamenti, l'applicazione del *reverse geocoding* dovrebbe fornire tratti stradali identificati da un intervallo di numeri civici. Se così è, selezionando un numero casuale all'interno del raggruppamento dei civici forniti dal *reverse geocoding* è possibile generare casualmente un indirizzo. In tutti gli altri casi, quando il programma restituisce solo il nome della via, sarà necessario l'intervento di un rilevatore per individuare un civico pertinente. Seguendo questo procedimento, la costruzione di una lista di indirizzi diventa una operazione praticabile e "veloce".

L'individuazione di un numero civico per la consegna del questionario non risolve completamente il problema dell'eleggibilità o meno dell'edificio, della famiglia, dell'individuo. Per arrivare a definire in modo univoco e standardizzato la famiglia da contattare è necessario – come avviene quando il programma non restituisce alcun intervallo di numerazione – fornire al rilevatore alcune regole da seguire (regole di *fieldwork*)⁹⁷. Ciò che comunque preme sottolineare è che, attraverso tale procedimento, l'individuazione di un edificio a cui recapitare il questionario diventa un'operazione possibile anche in assenza di una lista precostituita. Il primo scoglio nella costruzione di un campionamento areale è risolto.

⁹⁷ Vedremo in seguito quali sono state le regole dettate ai rilevatori che si sono occupati del progetto ISSP Italia (cfr. cap. 4.3).

3.2.2.2 Lo sviluppo verticale delle aree urbane

Rimane ancora da affrontare la questione della coerenza tra distribuzione delle unità campionarie e distribuzione della popolazione sul territorio oggetto d'indagine. Contrariamente a quello che succede usando i numeri di telefono, una scelta del tutto casuale dei punti geografici porta a sovrastimare le aree isolate. La probabilità di selezionare un particolare tratto stradale, infatti, è proporzionale alla dimensione del territorio circostante di cui il tratto stesso rappresenta la strada più vicina. Applicando la funzione di *reverse geocoding*, ad esempio nel comune di Roma, si ottiene un campione particolarmente sbilanciato verso le aree periferiche della città. Non inserire l'opzione che la scelta casuale delle unità campionarie rispetti la distribuzione della popolazione in base allo sviluppo verticale del territorio, significa, inevitabilmente, raccogliere dati (a livello socio-economico) distorti. Come è possibile tenere conto dello sviluppo verticale degli aggregati urbani?

Una prima soluzione potrebbe essere quella di suddividere il territorio in tante porzioni, tali da contenere un pari numero di abitanti. Successivamente, si potrebbe applicare un campionamento stratificato o a stadi che sfrutti il *reverse geocoding* quale tecnica per ottenere un set di indirizzi validi entro quelle porzioni di territorio. Questa soluzione, tuttavia, sottende la conoscenza a priori dell'esatta distribuzione degli abitanti sul territorio, elemento invece a noi sconosciuto e dal quale siamo partiti. In realtà, alcune informazioni a riguardo, con qualche difficoltà, le si potrebbero desumere:

- Le sezioni di censimento sono costruite con l'obiettivo di rendere omogeneo il lavoro dei rilevatori. La quantità di famiglie/persone entro ogni sezione non è perfettamente identico, ma, conoscendone l'entità, si può ottenere un campione con un ottimo grado di approssimazione. La difficoltà nel perseguire tale soluzione sta nella disponibilità, nel costo, nell'accuratezza e nel grado di aggiornamento della cartografia delle sezioni (*database geografico*).

- Le innovazioni tecnologiche in ambito geografico sono figlie di esigenze nate in ambito militare (aerospaziale) e/o attinenti al controllo del territorio (quello che in esso accade o può accadere). Per farsi un'idea di quanto sia utile e di quanto si sia investito nelle tecnologie GPS, si può esaminare il progetto governativo americano *LandScan* consultando il sito <http://www.ornl.gov/sci/landscan>. Si può verificare come per l'intero pianeta, esistano stime circa la distribuzione della popolazione presente per ogni km² all'interno delle 24 ore. Per alcune particolari aree degli Stati Uniti, ad esempio, la stima arriva ad un dettaglio di circa 90 m² e, in generale, sempre per il territorio americano, le mappe cartografiche sono presentate a dettaglio di circa 400 m². A cosa serve conoscere la numerosità della popolazione presente in un territorio entro le 24 ore ad un dettaglio così spinto? Pensiamo alla necessità di far fronte ad un disastro naturale, lo scoppio di una fabbrica chimica... Le informazioni georeferenziate rappresentano sicuramente una scommessa ed una opportunità in molti campi connessi all'attività di governo di una nazione. Utilizzando queste informazioni⁹⁸, come vedremo in seguito, alcuni istituti di ricerca americani hanno potuto pianificare i loro campioni con una granularità territoriale particolarmente minuta, più dettagliata rispetto a molte delle nostre sezioni di censimento.

In generale possiamo affermare che una mappa della distribuzione della popolazione sul territorio non è impossibile da trovare. Tuttavia, l'acquisizione può essere costosa ed è necessario avere una certa dimestichezza con gli strumenti informatici GIS per poterla utilizzare. È evidente che grazie a tali mappe è possibile tenere in debito conto il tessuto antropico del territorio e, in coerenza ad esso, sviluppare piani di campionamento attraverso il *reverse geocoding*.

⁹⁸ La disponibilità dei dati a dettaglio minuto è vincolata e sottoposta ad autorizzazioni specifiche.

Una seconda soluzione può essere quella di operare per via sperimentale utilizzando variabili *proxy* e metodi campionari per stimare la popolazione residente nelle varie porzioni di territorio. Nel progetto ISSP Italia si è optato per questa seconda soluzione. Una delle fonti più economiche e consistente (sebbene non completa) riguardante gli indirizzi di residenza della popolazione rimane l'elenco telefonico. È possibile utilizzare elenchi anche non aggiornati o, ancor meglio, effettuare dei *matching* tra vecchi e nuovi elenchi. In questo modo, si possono conoscere nuove vie/abitazioni e sopperire alla tendenza, sviluppatasi nel tempo, di non far comparire il proprio numero. Con poche centinaia di casi (indirizzi estratti casualmente) è alta la probabilità di ricostruire in modo appropriato la dislocazione delle famiglie sul territorio⁹⁹. Inoltre, cosa importante, questo procedimento evita la costruzione di partizioni territoriali a priori. Il procedimento di selezione degli indirizzi è piuttosto semplice e richiede di:

- a) Campionare casualmente “n” indirizzi dall'elenco telefonico
- b) Geocodificare gli indirizzi (restituzione della loro coordinata)
- c) Costruire delle poligonali (cerchi e/o quadrati) con centro le coordinate di cui al punto precedente¹⁰⁰
- d) Applicare il *reverse geocoding* su coordinate generate casualmente all'interno della poligonale di cui al punto precedente
- e) Estrarre casualmente un numero civico all'interno del tratto stradale identificato da due civici e generato con il *reverse geocoding*.

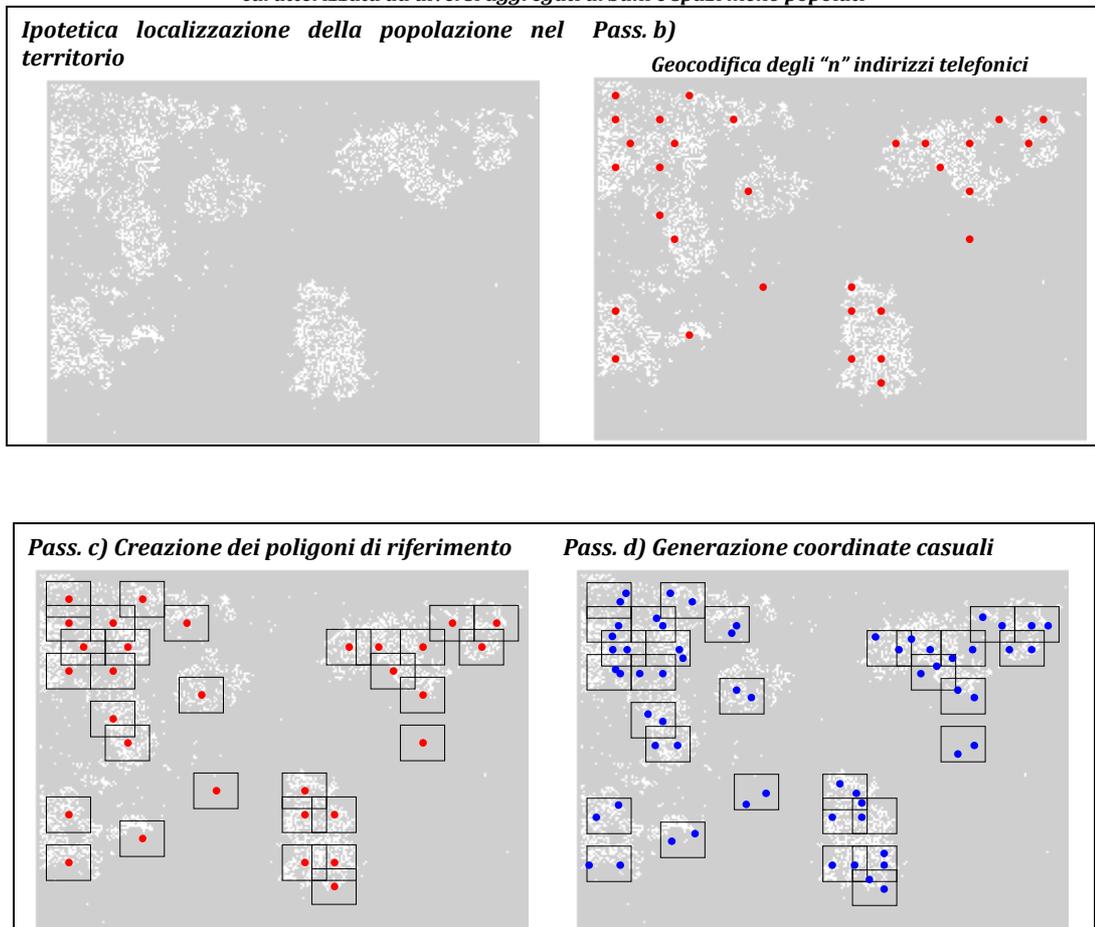
⁹⁹ In realtà con la geocodifica di tutti (o della maggior parte) gli indirizzi presenti nell'elenco telefonico si otterrebbe non solo una mappa più accurata della distribuzione della popolazione entro il comune, ma si avrebbe modo anche di valutare la qualità stessa del *database* georeferenziato che sta alla base del processo di geocodifica. La qualità della base informativa, come vedremo più avanti, è uno degli elementi sostanziali nella costruzione di piani di campionamento areali basati sulla funzione di *reverse geocoding*.

¹⁰⁰ La costruzione delle poligonali non deve avvenire necessariamente tramite la definizione di una mappa reale attraverso un programma GIS. Dato un punto, infatti, variando le coordinate di qualche millesimo (0,001 gradi), si possono ottenere i vertici di un territorio d forma quadrata avente 100 m circa di lato.

Tale procedimento è percorribile anche senza la disponibilità di strumenti GIS, utilizzando un po' di programmazione e le API (*Application Programming Interface*) di Google Maps.

È forse il caso di sottolineare ancora una volta che l'evoluzione tecnologica (il web) e la disponibilità dell'informazione risultano cruciali e indispensabili nell'applicazione di tale metodologia. La tecnologia, ritornando a quanto sviluppato nella prima parte della tesi, non costituisce assolutamente un impedimento all'evoluzione delle tecniche probabilistiche applicabili alla scienza sociale. Al contrario, ne possono amplificare e sviluppare i confini.

Figura 3.8: Il processo di reverse geocoding: creazione di una lista di indirizzi in una ipotetica porzione di territorio caratterizzata da diversi aggregati urbani e spazi meno popolati



3.2.3 Prime riflessioni sulla strategia di campionamento sviluppata

Come illustrato nel paragrafo precedente, sebbene non esista una lista completa degli indirizzi, attraverso il *reverse geocoding* e l'utilizzo di fonti dati parziali (elenco telefonico), anche in Italia è possibile ottenere un campione sperimentale attraverso una doppia procedura di campionamento. L'applicabilità della tecnica è legata all'ipotesi che le dismissioni telefoniche su rete fissa (l'elenco copre ancora circa il 65% delle famiglie) siano omogenee nei territori indagati e che, allo stesso modo, gli eventuali allacciamenti presso le nuove aree di urbanizzazione avvengano in misura omogenea. L'accumulo storico di informazioni sugli indirizzi della popolazione – possibili sia attraverso la consultazione di vecchi elenchi, sia utilizzando altre fonti, quali stradari – porta a una copertura piuttosto elevata della struttura geantropica del territorio.

Lo sviluppo e l'interesse crescente delle tecnologie geo-spaziali dovrebbe assicurare a questa tecnica un'applicabilità sempre maggiore nel prossimo futuro, grazie all'accrescimento della qualità dei dati presenti negli ambienti GIS (*database*). Recentemente (primi mesi del 2013), ad esempio, il produttore dei navigatori satellitari TOMTOM® ha reso noto i tempi medi che gli italiani trascorrono bloccati nel traffico con le proprie auto. L'analisi, svolta con milioni e milioni di informazioni che i navigatori satellitari trasmettono tra loro, di per sé può sembrare di poco conto, ma dà l'idea delle potenzialità insite in queste strumentazioni. Altri analoghi esempi sono riferibili alla possibilità di trasmettere la propria posizione con lo *smartphone*, così che amici e conoscenti possano verificare la nostra presenza nella zona in cui loro stessi transitano in quel momento. La disponibilità e la velocità d'accesso a queste informazioni rappresenta sicuramente una grande sfida per la ricerca sociale, più di quanto non lo siano le trasformazioni tecnologiche in sé.

I livelli di copertura e qualità delle informazioni georeferenziate in Italia sono ancora tutte da verificare. È certo che esistono (come negli USA) diversità tra territori e tipologie d'area (urbana vs rurale; Nord vs Sud), superabili comunque con un'adeguata organizzazione di controllo delle attività svolte sul campo dai rilevatori. Nei due anni di esperienza con l'indagine ISSP Italia (2011-12) si è potuto accertare la validità del metodo e la sua economicità ma, soprattutto, la possibilità di avere un pieno controllo del processo di campionamento da parte dei ricercatori.

Nello stesso periodo in cui l'indagine ISSP Italia era in fase di avvio, il dipartimento di geografia dell'Università dell'Iowa ha pubblicato un articolo nel quale si dà conto per la prima volta del medesimo metodo per sviluppare un'indagine pilota della *General Social Survey* nell'area metropolitana di Chicago (Kumar et al., 2011)

¹⁰¹. Nell'articolo viene citato il NORC¹⁰² come uno tra i principali fruitori-attori di esperienze di ricerca scientifica basate su campioni di indirizzi (ABS-DSF). Il NORC, che effettua le rilevazioni ISSP negli USA, è decisamente il centro di ricerca universitario che più di altri ha studiato/comparato, modificato e implementato metodologie su indirizzi postali quali alternative alla tecnica CATI-RDD¹⁰³.

Prima di approfondire la portata innovativa del *reverse geocoding*, ovvero l'impatto del campionamento areale nelle scienze sociali, è forse utile offrire una visione d'insieme del piano di campionamento utilizzato in Italia e le linee guida con cui è stato sviluppato. Ciò sarà utile anche per comprendere le differenze con l'unica proposta simile nota, quella sviluppata dal dipartimento di geografia dell'Università dell'Iowa.

¹⁰¹ Le differenze sostanziali tra il piano di campionamento utilizzato in Italia e quello americano saranno oggetto d'analisi in seguito (cfr. cap. 3.4). Quel che è certo, è che la strategia è sostanzialmente la stessa: utilizzo della funzione di *reverse geocoding* per ottenere un campione di indirizzi geo-casualizzato.

¹⁰² National Opinion Research Center dell'Università di Chicago.

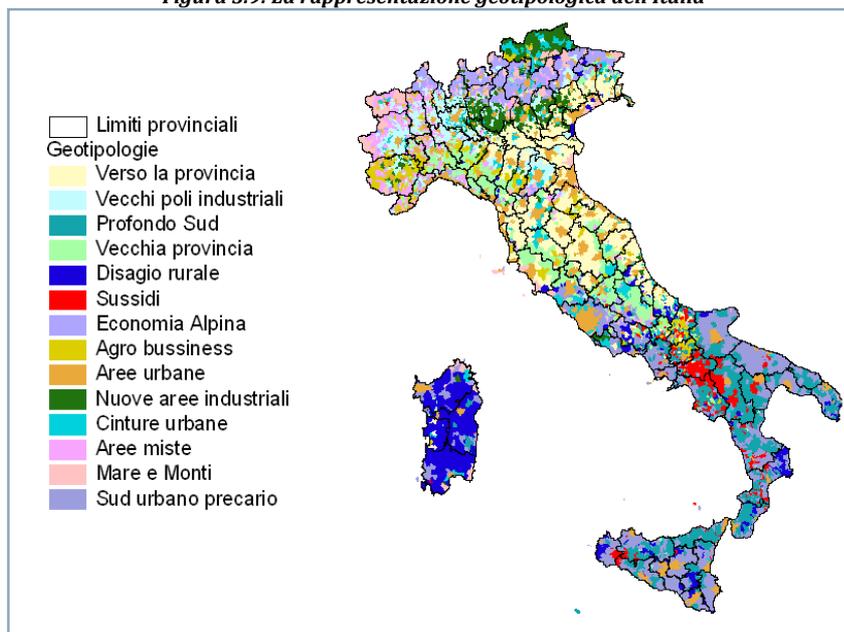
¹⁰³ Altre esperienze interessanti di utilizzo dell'ABS si possono trovare tra quelle proposte da RTI International (Research Triangle Park) del North Carolina e dall'Institute for Employment Research (DE). Molte sono pubblicate dall' AAPOR (American Association for Public Opinion Research).

3.3 *Dettagli e linee di sviluppo del piano di campionamento ISSP Italia*

3.3.1 Primo stadio: la stratificazione per geotipologie

La prima **suddivisione del territorio italiano** è avvenuta attraverso l'ausilio delle **geotipologie** che rappresenta un modo diverso di ripartire il territorio rispetto all'appartenenza del comune a determinate aree amministrative e/o strettamente dimensionali. Le caratteristiche socio-anagrafiche della popolazione residente, le fonti di reddito e le peculiarità produttive e commerciali del territorio vengono messe in rilievo in modo indipendente dalla posizione che il comune occupa nella geografia del paese¹⁰⁴. Come visibile nella Figura 3.9, le geotipologie manifestano comunque una sostanziale continuità geografica.

Figura 3.9: La rappresentazione geotipologica dell'Italia



¹⁰⁴ Le Geotipologie sono un prodotto di ricerca sviluppato dalla società Marker Srl di Venezia. Si tratta di un'analisi dei cluster non gerarchica, condotta con la tecnica delle Nubi Dinamiche, su dati prevalentemente di natura censuaria e inerenti i comuni italiani. Introdotta quale strumento di lettura della dinamica e della diffusione istantanea dei fenomeni di consumo nelle analisi di Geomarketing, le geotipologie sono state prodotte la prima volta nel 1994 e successivamente riviste in occasione del censimento 2001. I dati di base riguardano gli attributi demografici dei comuni (classe dimensionale, indicatori relativi alla ripartizioni per età e sesso della popolazione residente, dimensioni e tipologie delle famiglie), quelli morfologici (superficie, densità abitativa, tipologia degli aggregati urbani, classe altimetrica) e quelli - più consistenti - relativi alla struttura socio-economica del territorio (livelli di istruzione, caratteristiche e vestustà delle abitazioni, titoli di godimento delle stesse, livelli di occupazione per settori economici, condizioni professionali, popolazione attiva e non, ecc.).

L'introduzione delle geotipologie, al pari del metodo di definizione della lista degli indirizzi presso cui consegnare il questionario, rappresenta una novità nell'ambito del campionamento e per ISSP Italia. Esse disegnano una mappa della disomogeneità presente nel nostro paese e sono di fatto correlate ad alcuni fenomeni rilevanti per le indagini sociali ed economiche. Nell'introduzione all'articolo di Kumar et al. (2012), sono citati vari studi internazionali che avvalorano la tesi che il contesto geografico (socio-fisico) influenza nel lungo termine le condizioni di salute, le attitudini, l'ambiente sociale, i redditi e numerosi altri aspetti delle persone, così da avvalorare trasformazioni contestualizzate degli elementi significativi delle società. Partire dalle diversità socio-economiche dei comuni è sembrato anche per il progetto ISSP Italia il modo più adeguato per poter cogliere la variabilità degli argomenti d'indagine. L'obiettivo specifico è quello di rappresentare la massima variabilità possibile dei diversi fattori socio-economici con il minimo numero di classi, così da non imbattersi nella scelta di punti di campionamento tra loro troppo simili (ridondanza informativa). Tale principio, sebbene sviluppato a partire da conoscenze diverse e con metodi dissimili, è anche quello che vedremo in seguito suggerito dagli autori americani per la creazione di piani di campionamento areali ottimali (Kumar et al., 2012) (cfr. par.3.4.2).

3.3.2 Secondo stadio: i punti di campionamento (comuni)

L'indagine è stata pensata fin da subito come un'attività "porta a porta" attraverso l'ausilio di una rete di rilevatori. In tal senso, un primo punto fondamentale nella logica di sviluppo del piano di campionamento è stato quello di **limitare al massimo i punti di campionamento entro gli strati** (il numero dei comuni selezionati) perché ogni spostamento/selezione fa aumentare i costi. Le geotipologie, d'altro canto, non rappresentano la prima e l'unica stratificazione del campione, in quanto:

- non sono “contenitori” strettamente affini (necessariamente correlati) agli argomenti d’indagine, ma rappresentano delle condizioni strutturali omogenee del territorio di residenza della popolazione
- hanno consistenze diverse in termini di popolazione rappresentata. Quella meno consistente racchiude circa l’1% della popolazione (*agrobusiness*); quella più ampia raggiunge circa il 37% (*centri urbani*).

Figura 3.10: La composizione geotipologica dei comuni d’Italia

ID	GEOTIPO	Comuni nel geotipo	% pop. residente
9	Aree urbane	375	36.5%
14	Sud urbano precario	765	16.3%
2	Vecchi poli industriali	933	11.6%
1	Verso la provincia	968	10.3%
11	Cinture urbane	386	5.3%
10	Nuove aree industriali	676	5.3%
3	Profondo Sud	683	4.1%
5	Disagio rurale	460	2.6%
6	Sussidi	316	1.7%
13	Mare e monti (vacanze)	268	1.5%
4	Vecchia provincia	668	1.5%
7	Economia Alpina	576	1.3%
12	Aree miste	546	1.0%
8	Agrobusiness	459	0.9%

Da questo, la necessità di **limitare il danno connesso alla scelta di un numero ridotto (se non uguale) di comuni entro ogni strato**. Selezionare un numero di comuni particolarmente basso (ad esempio, e come limite, uno per strato) è rischioso perché potrebbe produrre stime con una distorsione proporzionale al peso dello strato sull’universo. Ecco il motivo per cui si è deciso che i territori più popolosi dovessero essere rappresentati da un numero di punti di campionamento (comuni) più elevato rispetto ai geotipi meno consistenti.

I comuni all’interno di ogni geotipologia sono stati estratti, con diversa numerosità, attraverso un campionamento in blocco. Il numero complessivo di comuni selezionati è stato di poco superiore a 60, una dimensione dettata prevalentemente dalle disponibilità economiche e definita, per un’insieme di ragioni, il limite inferiore al quale far riferimento.

3.3.3 Limitazione del danno, dei costi e le diversità geotipologiche: un approfondimento sulla rete dei rilevatori

Anche l'estrazione dei comuni rappresenta un elemento di qualità rispetto alla prassi che spesso vede la scelta dei punti di campionamento, relativi alle indagini dirette e promosse da diversi istituti di ricerca privati, dettata più dalla presenza in loco di un rilevatore piuttosto che da una vera e propria estrazione casuale dei siti. Rete di rilevatori, tra l'altro, frequentemente condivisa tra diverse agenzie di ricerca. Il fatto di non dover/poter presidiare l'intero territorio, sembra essere una regola consolidata e spacciata spesso per accettabile. D'altra parte anche le indagini ufficiali Istat ci hanno abituato alla pratica di un presidio del tutto parziale del territorio italiano. Nell'indagine sulle forze di lavoro, parte del programma statistico nazionale, cioè dell'insieme delle rilevazioni statistiche di interesse pubblico, il numero di comuni presidiati è 1.100 circa sugli 8.000 ed oltre presenti in Italia (Istat, 2006). A sostegno della validità dell'impianto metodologico che L'Istat adotta per l'Indagine sulle forze lavoro, è necessario comunque far osservare che 1.100 punti di presidio, 250.000 famiglie (oltre 600.000 interviste all'anno) e l'applicazione di un metodo di selezione delle stesse basato su criteri di rappresentatività dell'intera popolazione italiana e gestito tramite le anagrafi comunali è cosa ben diversa dal voler rappresentare con un campione particolarmente ridotto la disparità ambientale presente in Italia.

Sondare un campione casualmente distribuito sull'intero territorio italiano è ritenuto, da molti istituti di ricerca privati, praticamente impossibile da sostenere economicamente. La struttura dimensionale e geografica dei comuni italiani rappresenta sicuramente una complicazione nell'attività di ricerca. Altrettanto evidente, però, è la diversità e la concentrazione di strutture socio-demografiche ed economiche messe in luce dalle geotipologie. Aggiungerei una domanda: per quanto poco continuo

numericamente le aree di “disagio rurale” che si concentrano soprattutto in Sardegna (2.6% circa), chi ne garantisce il presidio? Concentrarsi sulle realtà più popolate, significa perdere una parte della variabilità nella composizione della struttura economica e sociale dell'Italia. Aspetto che si è voluto evitare nell'indagine ISSP selezionando casualmente da ogni geotipologia alcuni comuni (almeno uno). Ciò ha comportato la necessità di **costruire dal nulla una propria e pertinente rete di rilevatori**, selezionati valutando precisi profili tra i residenti nei comuni campione così da utilizzare appieno la loro conoscenza del territorio.

3.3.4 Terzo stadio: le famiglie contattate all'interno dei comuni sondati

La scelta, fatta al secondo stadio, l'estrazione di un numero di comuni pressoché proporzionale al peso dello strato, si coniuga bene con la necessità di comparare i risultati tra i diversi rilevatori. In questo senso, si è preferito **omogeneizzare gli sforzi, i costi e le remunerazioni tra gli operatori** (uno per comune) scegliendo di contattare un pari numero di famiglie (circa 100) entro ogni comune.

Intervistare entro ogni comune un numero uguale di persone, unito al fatto che al secondo stadio la selezione del numero dei comuni è avvenuta proporzionalmente alla dimensione demografica della geotipologia, ha permesso di ottenere un **campione autoponderato**. La **selezione casuale degli indirizzi** è avvenuta poi attraverso l'**applicazione del reverse geocoding**¹⁰⁵, metodo **che**, come visto precedentemente, **tiene conto dello sviluppo verticale degli aggregati urbani**.

3.3.5 Quarto stadio: i rispondenti

Sul frontespizio del questionario consegnato alla famiglia era riportata una data (giorno/mese) generata e abbinata casualmente agli stessi. Si

¹⁰⁵ Per ogni comune è stato estratto un numero di indirizzi superiore alle interviste previste per permettere di effettuare successivamente le sostituzioni necessarie nei casi di “mancato recapito” o di rifiuto.

chiedeva di far rispondere al questionario il **membro della famiglia con età superiore a 16 anni con la data di nascita più vicina rispetto a quella indicata sul questionario.**

3.4 Il campionamento areale per le scienze sociali: la proposta americana

3.4.1 L'impatto della tecnologia sulle prassi e sulle teorie del campionamento

Per comprendere la portata della tecnologia geo-spaziale nell'ambito delle scienze sociali e come essa influenzi l'ambito teorico e analitico-progettuale del campionamento, è il caso di ripercorrere a grandi linee quanto affrontato in questa terza parte della tesi.

Si è partiti affermando che la tecnica CATI è in profonda crisi e si è verificato che un'alternativa plausibile sono le indagini dirette o postali basate sugli indirizzi. In particolare, si è ribadito che lo sviluppo tecnologico (nell'organizzazione del lavoro, del web e nel comparto geo-informativo) consente oggi di articolare le indagini in modo da renderle efficaci, metodologicamente corrette ed economicamente paragonabili a quelle svolte con la tecnica CATI-RDD. Se con il *reverse geocoding* si supera addirittura la necessità di una lista dell'universo di riferimento, si comprende l'importanza dell'innovazione espressa dallo strumento geo-spaziale. Una trasformazione che ha spinto i ricercatori americani ad affermare che la tecnologia geo-informativa rappresenta una "nuova era" e che il processo di *reverse geocoding*, in particolar modo, accresce sia il ruolo del campionamento areale nelle indagini sociali, sia quello della ricerca multi-livello applicata alle scienze sociali. Un punto di vista decisamente entusiastico e in parte condivisibile. Approfondiamone i motivi.

Il primo articolo americano relativo all'applicazione del *reverse geocoding* quale metodo per la generazione degli indirizzi (Kumar et al., 2011), vede i ricercatori impegnati nel dar conto di come questa tecnica

permetta di superare la necessità di disporre a priori di una lista. Essi dividono la loro esperienza in quattro fasi:

1. la selezione dei siti campione da usare nel disegno campionario;
2. lo sviluppo della lista di indirizzi quale pseudo universo di campionamento;
3. la selezione degli indirizzi quale campione effettivo di lavoro;
4. la comparazione dei risultati di cui al punto 2. con i *file* DSF.

Le prime tre fasi rappresentano l'iter metodologico che sta alla base della proposta; il quarto è orientato a dar conto della qualità del percorso e poter affermare di aver superato a pieno titolo la necessità della lista iniziale.

L'articolo successivo (Kumar et al., 2012) si concentra sulla scelta ottimale dei siti di campionamento (fase 1) e sottolinea come l'approccio geo-spaziale migliori l'efficienza dei campionamenti areali. Viene inoltre messo in evidenza come l'evoluzione tecnologica e la concentrazione informativa nel web incentivino e rendano possibile lo sviluppo teorico delle tecniche di campionamento areale. I siti di campionamento, paragonabili alle unità di primo stadio nel campionamento tradizionale, sono scelti in modo efficiente. Un percorso di sviluppo teorico del tutto simile a quello avvenuto nella dottrina classica del campionamento probabilistico, che ha portato a sviluppare il campionamento a grappolo o stratificato quale tecnica di contenimento della numerosità campionaria (a parità di efficienza/errore) rispetto a quello semplice o in blocco (Diana e Salvan, 1993). Il poter presentare sotto un'altra veste le due colonne portanti della ricerca campionaria – la casualità nella scelta degli oggetti di osservazione e la potenziale copertura di tutto l'universo campionario –, nonché l'occasione di dimostrare possibili evoluzioni teoriche nella pratica campionaria, fa nascere nei ricercatori americani quell'espressione entusiastica a cui abbiamo fatto cenno. L'esperienza Italiana ISSP, in questo senso, si inserisce pienamente nell'ambito delle possibili generalizzazioni e

adattabilità di queste tecniche: parte da informazioni di base diverse, ma definisce un percorso di costruzione del campione del tutto in linea con l'impostazione proposta dal gruppo di ricerca americano.

A integrazione rispetto a quanto affermato nella prima parte della tesi, possiamo a questo punto asserire che l'evoluzione tecnologica ha effettivamente un impatto sulle tecniche di campionamento e sugli scenari della ricerca empirica. Offre il fianco a possibili derive ma, nel contempo, apre le porte a soluzioni metodologicamente corrette. Compito del ricercatore è quello di discernere i percorsi scientificamente accettabili.

3.4.1.1 La ricostruzione dell'universo di campionamento

Il primo dei due articoli citati (Kumar et al., 2011), introduce due sostanziali novità tecnologiche nella predisposizione dei piani di campionamento per le scienze sociali: le informazioni geo-spaziali ed il *web-mining*¹⁰⁶. Si tratta dell'uso di Google Maps quale base dati geografica per l'estrazione degli indirizzi e delle pagine bianche e gialle presenti in internet (elenco degli utenti telefonici, rispettivamente aziende e famiglie) quale step successivo per verificare e "filtrare" le risultanze di Google Maps. Vediamo nello specifico i passaggi proposti.

Step 1) Selezione delle micro-aree di campionamento

Dato un territorio, nel caso specifico l'area metropolitana di Chicago, è necessario innanzitutto individuare i punti (micro-aree) di campionamento. Il progetto governativo LandScan mette a disposizione delle mappe informative sulla popolazione ad "alta risoluzione" (cfr. par. 3.2.2.2). Quella utilizzata dai ricercatori americani vede una "granularità" di circa 400 m²: ogni "pixel" della mappa, l'unità di base che definisce uno spazio territoriale, rappresenta cioè un'area di circa 400 m². Considerando i dati censuari e il *database* LandScan, i ricercatori hanno sviluppato quattro

¹⁰⁶ Parlare di *web-mining* in questo esperimento specifico è forse eccessivo. Il *web-mining*, più propriamente, è l'utilizzo di tecniche specifiche di *data-mining* applicate al web. Nell'articolo citato, invece, si confrontano gli indirizzi provenienti dal *reverse geocoding* con i dati delle pagine gialle e bianche forniti via web.

metodi distinti per definire le unità di primo stadio. In due casi le aree selezionate coincidono con le aree di censimento (blocchi e tratti censuari); negli altri due, il territorio è suddiviso in spazi pari a 1 pixel (~400 m²), aree tratte dalla suddivisione territoriale offerta da LandScan. La numerosità totale dei punti di campionamento è risultata pari a 951.

Figura 3.11: Campionamenti adottati per le unità territoriali a Chicago (Kumar et al., 2011)

Area selezionata (unità territoriale di base)	Tipo di campionamento	
	Non Ottimizzato	Ottimizzato
Tratti e blocchi censuari	Campionamento sistematico delle unità ordinate in base ad alcune caratteristiche socio-economiche d'interesse. (tecnica applicata per la General Social Survey)	Definizione di un indice composito per le unità di base (che tiene conto della concentrazione della popolazione residente e delle caratteristiche di status socio-economico) e selezione casuale di un'insieme di unità in grado di catturare la massima variabilità dell'indice
1 pixel (~400 m ²)	Campionamento casuale semplice, tale da evitare l'autocorrelazione spaziale delle partizioni territoriali	Selezione casuale di un'insieme di unità in grado di catturare la massima variabilità dell'indice composito generato (vedi sopra), evitando l'autocorrelazione spaziale delle partizioni territoriali

Step 2) Generazione di uno pseudo-universo di campionamento (applicazione del *reverse geocoding*)

Per ogni punto dello step 1) è generata una coordinata casuale (appartenente allo spazio di riferimento) alla quale viene applicato il *reverse geocoding* di Google Maps. Il risultato (l'indirizzo) è confrontato con i *database* telefonici per verificare l'esistenza di almeno una unità abitativa entro l'edificio. La procedura è ripetuta fino al raggiungimento di un congruo universo campionario (20 indirizzi per ogni punto/area di campionamento). In totale, lo pseudo-universo di campionamento si compone di 19.020 indirizzi. Poiché le micro-aree avrebbero potuto essere definite da una bassa densità di popolazione o a esclusiva caratterizzazione commerciale, dunque non in grado di generare 20 indirizzi residenziali, per questi casi il processo è stato ripetuto aumentando l'ampiezza della micro-area con incrementi pari a 100 m (0,001 gradi).

Lo pseudo-universo di riferimento diventa una struttura “a macchie di leopardo” con una distribuzione più o meno omogenea di punti (indirizzi) entro micro-aree.

Step 3) Selezione casuale degli indirizzi (ultimo stadio del campionamento)

Partendo dall'universo di riferimento costruito nei primi due step, è stato scelto casualmente un numero di indirizzi sufficiente affinché il campione finale possa raggiungere la numerosità desiderata. In questa operazione il tasso di risposta atteso gioca un ruolo dominante. Nel caso in esame, dei 20 indirizzi per micro-area ne sono stati selezionati 14, ipotizzando un tasso di successo cautelativo pari al $10\% \pm 2\%$.

Step 4) Processo di validazione della lista degli indirizzi

Al fine di controllare la validità del metodo – verificare la bontà dell'informazione dedotta tramite i *database* geografici (Google Maps) –, i ricercatori hanno selezionato, tra i 20 di ogni micro-area, un solo indirizzo (951 in totale). Tali indirizzi sono stati confrontati con le anagrafiche presenti nel *file* DSF fornito da USPS (servizio postale americano). Nel 98% dei casi le anagrafiche dedotte dal *database* geografico (Google Maps) hanno trovato corrispondenza in quelle USPS. Per il rimanente 2% si sono riscontrati errori imputabili sia alla qualità del *database* DFS che a quella del *database* geografico di Google Maps. La sostanziale corrispondenza tra i due *database* (98%) conferma l'ipotesi dei ricercatori, ovvero la “maturità” delle informazioni geografiche presenti nel web. Si tratta un passaggio fondamentale in quanto consente un ulteriore sviluppo pratico-teorico. Attraverso la tecnica del *reverse geoconding*, come abbiamo già detto, è possibile ricomporre l'universo di riferimento. Oltre a ciò, grazie all'ausilio degli strumenti GIS è possibile calcolare le correlazioni spaziali dei territori caratterizzati dalle variabili censuarie o di altra natura. In questo modo è possibile far “trasmigrare” le informazioni da una macro-area ad una

particolare micro-area correlata spazialmente e quindi costruire piani di campionamento più efficienti (eliminazione della ridondanza informativa).

Prima di approfondire il tema dello sviluppo teorico dei campionamenti areali a cui abbiamo appena ora fatto cenno, oggetto del secondo articolo dei ricercatori americani (Kumar et al. 2012), è utile analizzare alcuni aspetti sostanziali dell'esperimento proposto per l'area di Chicago.

3.4.1.2 Sull'indice di bontà dell'esperimento e sul significato ultimo dell'esperimento proposto

Il processo di validazione della lista di indirizzi (step 4) consiste nel ripercorrere al contrario l'esperimento di geocodifica applicato ai *file* DSF. Con la geocodifica si parte dagli indirizzi presenti nei DSF e si tenta di posizionarli nel *database* geo-spaziale. In questo caso, invece, il dato proveniente dal *database* geo-spaziale (l'indirizzo fornito dalla funzione di *reverse geocoding*) viene ricercato quale indirizzo nel *database* DSF. L'indice di concordanza tra le fonti ottenuto dai ricercatori americani (98%) conferma, dal loro punto di vista, la qualità delle fonti e permette di comprovare l'attinenza funzionale del metodo di *reverse geocoding*. È importante ricordare, però, che la tesi vale se non si trascura l'importanza del passaggio "filtro" del *web-mining* e se si ha ben presente che l'indice non rappresenta il grado di copertura del *database* geografico. Infatti, per calcolare un indice della capacità della funzione di *reverse geocoding* di fornire risposte adeguate si dovrebbe tener conto non solo di eventuali problemi di copertura (abitazioni non censite nel *database* geografico sottostante, come strade abitate e "sconosciute" al sistema di navigazione), ma anche di tutti i "falsi positivi" (civici inesistenti) che la funzione stessa può generare. Facciamo un esempio. Se nel *database* geografico non esiste una determinata strada, qualsiasi punto del territorio che dovrebbe far riferimento a un indirizzo di quella via, attraverso la funzione di *reverse geocoding* viene mappato su un'altra via (la più prossima, tra quelle presenti nel *database* geografico, rispetto al punto individuato). In questo

caso, supponendo che il *file* DSF presenti una copertura totale, l'esperimento condotto dagli americani porterebbe ad un indice pari al 100%, mentre sappiamo che le abitazioni presenti nella via mancante avrebbero probabilità zero d'essere estratte. La qualità del *database* geografico, quindi, può essere valutata solo in ragione della capacità di offrire una soluzione in termini di geocodifica (posizionamento di un indirizzo) a fronte dell'utilizzo di un'anagrafe completa. I dati offerti da altri studi indicano che solo l'83,3% degli indirizzi presenti nei DSF *files* (che a loro volta offrono una copertura media del 92,3%) sono geocodificabili con precisione (cfr. par. 3.1.4.2). Di conseguenza, la copertura offerta dal *database* geografico, al netto dei "falsi positivi", dovrebbe essere non inferiore al 77% ($92,3\% \cdot 83,3\%$). Ipotizzando che parte degli indirizzi non presenti nei *files* DSF siano comunque presenti nei *database* geografici, ovvero che una parte degli indirizzi presenti nei DSF necessiti di qualche aggiustamento affinché il toponimo sia riconosciuto in modo preciso, ci si può spingere a dire che l'indice di copertura offerto dai *database* geografici è presumibilmente pari alla qualità dei *database* DSF, considerata ottima. È questo il riscontro ottenuto dai ricercatori americani circa gli errori analizzati in fase di *matching* delle anagrafiche.

In sostanza, la funzione di *reverse geocoding* può offrire, a seconda dei territori, risultati migliori o peggiori, in termini di copertura, rispetto a quelli dei DSF *files*. Le informazioni che oggi si possono ricavare da questi due *database* sono qualitativamente simili in contesti presidiati. A livello prospettico sarà interessante verificare quanto il processo di aggiornamento e accumulazione delle informazioni coinvolgerà i *files* DSF rispetto ai *database* geografici cioè, quanto si sedimenterà la concentrazione geoinformativa oggi condivisibile in internet. In questo senso, potremmo chiederci: quale prospettiva ha l'ipotesi e l'investimento di Google nel predisporre una tecnologia in grado di offrire un'auto che guida da sola, che ci accompagni in sicurezza nell'esatto punto in cui vogliamo essere trasportati?

Sebbene si possa condividere l'idea che le fonti georeferenziate rappresentino fin da subito uno strumento utile nella ricerca sociale, risulta invece difficile avvalorare l'ipotesi che la funzione di *reverse geocoding* rappresenti, oggi e da sola, uno strumento operativo per predisporre delle indagini postali a costi contenuti. Con l'attuale qualità/completezza informativa dei *database* georeferenzati, il passaggio al *web-mining* rimane un passaggio cruciale. Infatti, è ancora concreta la necessità di eliminare dalle soluzioni offerte dal *reverse geocoding* i "falsi positivi" (numeri civici inesistenti) e gli indirizzi ineleggibili (edifici privi di abitazioni). Da questo punto di vista l'esperimento americano è migliorabile laddove utilizza le pagine bianche e gialle presenti in internet come "setaccio" informativo, riconducendo l'esperimento campionario ad una mera estrazione di indirizzi dagli elenchi telefonici. Potremmo definirlo un peccato veniale, funzionale ad accreditare due ipotesi: il progressivo processo di concentrazione informativa promosso dalla tecnologia web (che porterà verosimilmente ad incrementare le fonti dati utili per il *web-mining* e la qualità stessa del *reverse geocoding*), e la possibilità di creare campioni pur in assenza di una conoscenza specifica dell'universo di riferimento (la lista).

Nel recente passato la lista dell'universo era rappresentata dall'elenco telefonico, corretto eventualmente con la tecnica RDD per sopperire a problemi di copertura. Successivamente si sono usati i *files* DSF e sviluppate tecniche per la perlustrazione del territorio al fine di verificare la piena copertura delle aree indagate. Oggi è possibile cominciare ad utilizzare le informazioni riferibili ai dati geo-spaziali: una lista virtuale pressoché completa relativa all'intera geografia del territorio. In sostanza, la tecnica del *reverse geocoding* può essere considerata a tutti gli effetti una metodologia per l'autogenerazione contestuale dell'universo di campionamento, esattamente quello che ha rappresentato l'RDD nella tecnica CATI su rete fissa. Un passaggio tecnico-metodologico di grande importanza operativa.

In conclusione, come già ribadito, il *reverse geocoding* si basa su un importante insieme di informazioni e rappresenta uno strumento adeguato per predisporre indagini dirette/personali nelle quali un apporto ulteriore all'identificazione della famiglia a cui recapitare il questionario può essere fornito da un "postino intelligente", un soggetto in grado di annullare gli eventuali indirizzi non pertinenti (è il caso dell'ISSP Italia). Al contrario, rimane ancora da dimostrare quanto esso possa, nell'immediato e da solo, sostituire completamente una generica anagrafe (come i DSF) e quindi permettere di effettuare indagini postali dove la consegna del questionario avviene tramite il servizio di recapito tradizionale. È comunque incontrovertibile che la tecnologia geo-spaziale e la concentrazione informativa in rete offrono e offriranno alle scienze sociali diverse sfide; sia dal punto di vista metodologico (tecniche di campionamento e di raccolta dati) che di analisi dei fenomeni.

3.4.1.3 Sulla problematica dello sviluppo verticale delle realtà urbane

Circa la questione dello sviluppo verticale delle realtà urbane e della conseguente necessità di rendere coerente il campione estratto con la struttura antropica del territorio, la proposta americana ne fa un parziale cenno solo in ragione del contenuto informativo connesso all'indice unitario¹⁰⁷ che descrive la diversità tra le varie micro-aree prese in esame (unità territoriali di base). Il campione sviluppato in ISSP Italia rappresenta, da questo punto di vista, un metodo per generalizzare la proposta campionaria statunitense da applicarsi nei casi limite in cui:

- l'indice unitario di descrizione delle micro-aree esprime soprattutto la diversa densità di popolazione (mentre si presume o è noto che anche altre componenti siano fonte di varianza per il fenomeno indagato), cioè quando è utile rivedere la costruzione stessa dell'indice unitario scorporando la descrizione dei territori dalla componente relativa di densità;

¹⁰⁷ Indicatore sintetico costruito attraverso analisi multidimensionali.

- le unità territoriali di base sono spazi ampi e/o particolarmente disomogenei al loro interno in termini di sviluppo abitativo e tale disomogeneità può essere ritenuta causa di una distorsione sulla stima.

È anche il caso di mettere in luce che l'uso del *reverse geocoding* ha delle implicazioni economico-operative rilevanti (spostamenti dei rilevatori). Operare in aree ristrette (come nel caso dei ricercatori statunitensi) o tenere conto dello sviluppo verticale delle abitazioni (come nel caso ISSP Italia) è quindi un passaggio operativo non indifferente. Nella metodologia applicata a ISSP Italia, per ricordare i due modi di procedere, la componente socio-economica e demografica che dà origine alle geotipologie è trattata disgiuntamente dalla densità di popolazione entro il comune, individuato al secondo stadio del campionamento. Si è adottata, infatti, una tecnica per la simulazione della densità abitativa nel territorio/comune selezionato (elenchi telefonici) e le geotipologie per la stratificazione del territorio. Una scelta di metodo obbligata dalla disponibilità delle informazioni di base che in ogni caso non incide sul carattere sperimentale del campionamento, ma, semmai, come si vedrà a breve, sull'efficienza/ottimalità del campione.

3.4.2 Lo sviluppo teorico dei campionamenti areali

In una seconda trattazione, i ricercatori americani (Kumar et al., 2012) affrontano il tema dell'individuazione ottimale dei punti/aree di campionamento su di un territorio a livello generale e avulso dalla tecnica di ricostruzione (*reverse geocoding*) o di conoscenza diretta dell'universo (applicazione di tecniche di campionamento basate su anagrafiche). L'attenzione si concentra sulla teoria del campionamento areale e sulla prospettiva di utilizzo congiunto delle fonti in linea con lo sviluppo tecnologico e teorico in atto.

3.4.2.1 La selezione delle unità campionarie

La teoria classica del campionamento ci dice che il processo di stratificazione dell'universo attraverso l'utilizzo di variabili/informazioni correlate al fenomeno indagato può migliorare l'efficienza del campione (Diana e Salvan, 1993). Inoltre, qualora siano note varianze e consistenze di ogni strato, seguendo le indicazioni di Neyman è possibile calcolare il numero minimo di casi campionari in grado di cogliere, a parità d'errore delle stime, la massima variabilità del fenomeno indagato (ripartizione ottima del campione stratificato). L'ausilio di variabili correlate al fenomeno indagato permette, in sostanza, di ottimizzare il campione, cioè contenere la numerosità a parità d'errore (o comprimere l'errore di stima a parità di casi).

Come applicare un'analoga strategia nel campionamento areale? Gli autori statunitensi affermano che ottimizzare un campionamento areale significa effettuare un'adeguata selezione di micro-aree che tenga conto della correlazione spaziale esistente tra i "granuli" (micro-aree) che costituiscono il territorio oggetto d'indagine. Nel campionamento areale, infatti, sono obiettivi fondamentali del processo di selezione sia il raggiungimento di una adeguata numerosità campionaria (connessa al livello di errore accettabile), sia la distribuzione spaziale degli oggetti indagati (collegabile al problema della ridondanza informativa). Il processo di estrazione dei punti di campionamento deve tendere a massimizzare la semi-varianza¹⁰⁸ espressa dal territorio indagato ed escludere le micro-aree spazialmente correlate a quelle che mano a mano vengono estratte. Così facendo si elimina la ridondanza informativa dovuta alla correlazione spaziale assicurando adeguatamente sia la "copertura" della variabilità espressa dall'universo di riferimento sia la dispersione dei punti di campionamento nello spazio fisico indagato. Per esemplificare, paragoniamo il dominio territoriale alle tessere di un puzzle: è come se, per

¹⁰⁸ La varianza (o semi-varianza) fa riferimento alla disomogeneità socio-economica (o semplicemente di densità abitativa) della popolazione residente nelle micro-aree.

ogni tessera estratta e inserita nel campione finale in base alla sua “diversità” rispetto alle altre (“tessera” che esprime la semi-varianza più elevata) si togliessero dal tavolo anche tutte quelle particolarmente “simili e vicine” (correlate spazialmente) a quella scelta. Ciò che si ottiene al termine del processo (il campione finale), è un insieme di tessere decisamente diverse tra loro e (inizialmente) ben distribuite sul tavolo. Di volta in volta e senza reinserimento, è individuata la micro-area che presenta la semi-varianza più elevata (che rappresenta il contributo informativo più cospicuo). Ad ogni estrazione, il calcolo della semi-varianza relativa ad ogni micro-area è quindi emendato dall’effetto di correlazione spaziale dettato dalle unità già selezionate. Un po’ quello che succede con una tecnica di analisi di regressione multipla applicando il metodo *forward selection* (Fabbris, 1983).

Dopo aver sviluppato la parte matematico-statistica per la selezione delle unità (micro-aree), l’articolo si occupa del processo di inferenza (stima della media e della varianza campionaria) e di come valutare l’incremento d’efficienza campionaria ottenuta. È una trattazione teorica particolarmente corposa integrata con una applicazione pratica su di un territorio specifico: di fatto l’approfondimento del campionamento realizzato con uno dei quattro metodi proposti nel primo articolo precedentemente illustrato (Kumar et al., 2011).

A detta degli stessi autori, il procedimento descrive un metodo ottimale per campionamenti su territori circoscritti. Nel caso si debba operare sull’intera nazione, essi propongono dei campionamenti a stadi, per i quali la tecnica suddetta dovrebbe rappresentare il secondo stadio (individuazione della famiglia da contattare). Come avvenuto per ISSP Italia, infatti, suggeriscono che le unità di primo stadio siano scelte in base a criteri di proporzionalità su aggregati territoriali omogenei (stratificazione del territorio nazionale). Tali *cluster* multidimensionali e spazialmente omogenei possono essere costruiti utilizzando le caratteristiche socio-economiche censuarie (*cluster* geotipologici) o, più semplicemente, le

suddivisioni topografiche, amministrative, ambientali e climatiche (come le nostre province, regioni, aree Nielsen); ovviamente quando è ipotizzabile, come spesso accade, una loro influenza sul fenomeno indagato. Tale appunto inserito nell'articolo, che ricordiamo essere stato pubblicato successivamente alla predisposizione dell'indagine italiana, avvalga ancor più la sovrapposibilità tra la tecnica campionaria usata in ISSP Italia e quella proposta dai ricercatori americani.

3.4.2.2 La propagazione delle informazioni censuarie a livello micro-areale

Un altro aspetto decisamente importante analizzato nel *paper* americano riguarda le tecniche di propagazione, al piano delle micro-aree, delle informazioni di tipo socio-economico e relative ad ampi dettagli territoriali e, in particolare, l'importanza dei dati micro-areali offerti dal progetto LandScan.

È noto che la conoscenza a priori di informazioni correlate al fenomeno d'interesse e riguardanti il territorio oggetto d'indagine può migliorare l'efficienza di un campionamento. Tali dati, spesso provenienti da fonti istituzionali, difficilmente propongono un dettaglio territoriale spinto (per micro-aree). In questo senso, diventa fondamentale comprendere se gli strumenti e/o la metodologia stessa di campionamento adottati non escludano la possibilità di attingere a questo patrimonio di conoscenze. La tecnologia GIS oggi disponibile e la conoscenza della concentrazione della popolazione per micro-aree, consentono l'utilizzo di informazioni non strettamente geo-caratterizzate o distinte per dettagli territoriali diversi rispetto alle micro-aree utilizzate nel campionamento. È fondamentale, però, definire un modello di propagazione dell'informazione che può essere una semplice distribuzione uniforme rispetto all'intensità di un altro fenomeno (quale la concentrazione/distribuzione della popolazione residente nelle micro-aree). In questo caso, ad esempio, la quantità di anziani nota a livello comunale può essere ripartita proporzionalmente alla popolazione residente nelle micro-aree. È possibile adottare anche modelli di propagazione più sofisticati, come una

regressione lineare semplice basata sulla dimensione (spaziale o in termini di abitanti) delle micro-aree sul quale viene riversato il dato. Ciò che caratterizza la tecnologia GIS, rispetto a queste pratiche statistiche “classiche”, è la possibilità di tener conto del posizionamento spaziale degli oggetti e, di conseguenza, delle eventuali correlazioni spaziali tra i fenomeni, così da mettere in relazione sia il volume demografico delle micro-aree sia le direttrici di diffusione del fenomeno nello spazio circostante. Come più volte sottolineato, l’analisi spaziale dei fenomeni e la possibilità di caratterizzare le micro-aree in ragione di conosciute variabili riguardanti altre dimensioni territoriali, arricchisce la classica analisi dei dati con la dislocazione spaziale degli oggetti. Sostanziale, per i ricercatori americani, è almeno la conoscenza delle taglie demografiche delle micro-aree prese in esame, cioè le informazioni provenienti dal progetto LandScan.

Per quanto ancora grezza e perfettibile possa essere la metodologia sperimentata con ISSP Italia – campionamento su elenchi telefonici e geocodifica per la stima dello sviluppo verticale degli aggregati urbani –, si conferma un valido strumento per il processo di inferenza qualora si vogliano analizzare dinamiche a livello micro-areale. In questo senso la proposta italiana si può intendere come una possibile generalizzazione di quella americana, in linea con le risorse informative ed economiche a disposizione.

In conclusione, i due articoli americani (Kumar et al., 2011 e 2012) confortano la scelta italiana di dotarsi di una propria rete di rilevatori, invece di replicare in proporzioni ridotte la tecnica campionaria proposta dall’Istat per l’indagine sulle forze lavoro e così spesso simulata da diversi istituti di ricerca. Generalmente i campionamenti non assumono le caratteristiche socio-economiche del territorio quali “ambiti” preliminari per la valutazione dei punti “ottimi” da selezionare rischiando, quindi, una possibile elevata ridondanza informativa.

3.4.3 Sintesi e spunti di riflessione sullo sviluppo della tecnologia geo-informativa

Il campionamento areale basato sulle informazioni geo-spaziali rappresenta una sostanziale novità nel panorama della ricerca sociale. L'attenzione dei ricercatori verso questi argomenti va inserita comunque in un contesto più generale di evoluzione tecnologica orientata alla geo-informazione. È un ambito al quale appartengono anche le esperienze di geo-marketing e di analisi geo-ambientale. Il percorso evolutivo-strumentale ha quindi origini lontane, a riprova dell'influenza che le tecnologie e le conoscenze hanno sull'ambiente sociale e culturale. Un percorso autopropulsivo così come l'aveva immaginato Arthur Young già due secoli fa (cfr. nota 61).

La questione fondamentale per la ricerca sociale applicata è come perseguire in modo costruttivo il percorso evolutivo. Servono strumenti in grado di coniugare lo sviluppo teorico e quello tecnologico, formalizzando progetti "infrastrutturali" in grado di favorire l'applicazione stessa della teoria. Volendo presentare un paragone, sarebbe paradossale che alla scoperta del ferro non seguisse una rivoluzione sociale perché nessuno coglie l'utilità di realizzare un incudine ed un martello con cui forgiare gli strumenti che con questo materiale si possono costruire! Come si è messo in evidenza nella terza parte della tesi, la praticabilità del campionamento areale ottimale è condizionata alla conoscenza della geoantropia del territorio. LandScan, in questo senso, risulta un progetto ineguagliabile in termini di utilità. Una fonte dati di cui l'Italia parrebbe non disporre, a cui non partecipa e sulla quale non sta investendo in termini progettuali.

Le informazioni sulla popolazione residente a livello micro-geografico permetterebbero il raccordo tra fonti (ad esempio quelle censuarie a livello di sezioni di censimento) e amplificherebbero la possibilità di sviluppare piani di campionamento sperimentali efficienti. Qualcuno potrebbe dire che l'aver formalizzato questo percorso non aggiunge nulla di nuovo. Invece,

prospettando l'opportunità di realizzare indagini attraverso azioni mirate sui territori con l'ausilio di informazioni già presenti nel web, tale apporto teorico apre la strada ad una nuova "geografia campionaria"; un mondo tutto da esplorare dal punto di vista tecnico-metodologico.

Si è detto che lo studio spaziale dei risultati d'indagine o delle informazioni di partenza apre ulteriori scenari nella ricerca scientifica sociale. È ad esempio possibile condurre analisi sulla distribuzione micro-areale della popolazione per caratteristiche etniche, sociali ed economiche e studiare le dinamiche e le concentrazioni di alcuni fenomeni utili per attivare specifiche politiche sociali e di governo del territorio. Analogamente, con questi strumenti oggi si semplifica anche la costruzione di campioni metodologicamente corretti/scientifici.

In sintesi e per concludere, le informazioni georeferenziate offrono nuove possibilità interpretative dei fenomeni e amplificano le opportunità d'indagine perché si arricchiscono del concetto di spazialità dell'informazione. Il campionamento areale, nello specifico permette di superare i confini amministrativi dei territori e considerarli – come di fatto sono – spazi contigui rispetto alle sovrastrutture amministrative con le quali solitamente si opera. Si pensi alla possibilità di svolgere un'indagine campionaria in una determinata area (anche lo stesso comune), stratificando il territorio sulla base della densità di popolazione e aggregando territori amministrativi contigui sulla base dell'omogeneità socio-economica (geotipologie) e/o dei carichi di lavoro (distanze/percorsi) dei rilevatori, così da omogeneizzare le attività sul campo.

Il *reverse geocoding* (tecnica di reperimento degli indirizzi), permette:

- la realizzazione di piani di campionamento pienamente sperimentali (copertura e processo di selezione casuale)
- il contenimento dei costi

- il pieno controllo delle linee di campionamento da parte dei ricercatori
- il superamento della necessità di avere una lista completa relativa all'universo di riferimento, in quanto essa rappresenta una tecnica per la generazione contestuale dell'universo di campionamento.

Non ultimo, come abbiamo potuto verificare per ISSP Italia, essa può rappresentare una metodologia generalizzabile in vari contesti informativi.

Tutto questo non scardina e non annichilisce ciò che la dottrina e l'esigenza di ricerca ci insegna e ci impegna oggi. Si tratta "solo" di un ulteriore avanzamento in linea con la trasformazione in atto.

4 – Il fieldwork e alcuni risultati ottenuti

4.1 Premessa

Nella prima parte della tesi si è dato corpo all'ipotesi che la tecnologia, soprattutto nelle indagini dirette, è destinata (o dovrebbe essere destinata¹⁰⁹) a favorire sempre più l'avvicinamento tra ricercatori, gruppo di rilevatori e soggetti appartenenti al campione. Si è anche detto che uno degli attuali problemi della ricerca è l'espulsione di tutti quegli elementi organizzativi che caratterizzano la connessione tra le diverse fasi del processo produttivo; specie quando lo si vorrebbe altamente competitivo e dunque trasformato da artigianale a industriale-specializzato; ciò in ragione del sistema di pressione "ambientale" dettato dall'avanzamento tecnologico e dalla tipica evoluzione dei sistemi economici. Si è accennato alla distinzione tra le indagini *in-line* e *off-line*, dove le prime vedono confluire in tempo reale le informazioni di processo e di prodotto (i dati) presso i *database* aziendali, così da aumentare l'efficacia dell'azione di governo del processo e assicurare tempi ridotti nella predisposizione delle fasi successive. Si è visto che, data la stretta relazione esistente tra qualità del dato e processo di produzione, l'organizzazione complessiva del lavoro – attività umane, metodologie, strumenti e tecnologie – va studiata e seguita con grande cura. Spesso invece è data per scontata, pensata come ormai standardizzata e trascurata dal punto di vista dell'impatto sulla qualità del dato acquisito.

Partendo da queste premesse, per l'indagine ISSP Italia si è deciso, da un lato di adottare quanti più strumenti informatici possibili al fine di archiviare, gestire e rendere veloci i passaggi informativi tra i diversi attori della ricerca, dall'altro di curare adeguatamente l'organizzazione del lavoro a partire dalla selezione dei rilevatori. Una lista degli strumenti adottati nell'indagine ed alcune scelte operative sono già state descritte nella seconda parte della tesi. Nella Figura 4.1 sono riportati gli strumenti

¹⁰⁹ Sull'opportunità che la tecnologia diventi uno strumento di coesione tra gli attori della filiera produttiva dell'informazione, è da consigliare la lettura dell'articolo di Ermelinda Boccuzzi (1985).

raccolta dati e sul mantenimento del carattere sperimentale della ricerca. Data la rilevanza del tema, si è pensato di concludere, poi, fornendo alcuni risultati relativamente all'uso del CAWI quale strumento di partecipazione (risposta) all'indagine stessa. Ciò che è riportato in seguito, fa riferimento ai dati e alle procedure relativi all'ultima rilevazione ISSP Italia sviluppata e conclusa nel 2012 (l'indagine riguardante i moduli ambiente e salute).

4.2 I rilevatori

4.2.1 La scheda di consegna e il ruolo dei rilevatori

L'indagine ISSP Italia può essere definita una rilevazione a tecnica ibrida: semi-postale e CAWI. Semi-postale perché la consegna della busta contenente il questionario (e vari altri materiali) è avvenuta tramite un rilevatore direttamente a casa della famiglia selezionata (terzo stadio del campionamento). Ibrida, in quanto la restituzione del questionario poteva avvenire attraverso spedizione postale (con busta preaffrancata messa a disposizione allo scopo) o via internet (CAWI). L'indagine è stata anche definita come diretta ma non personale. Al rilevatore, infatti, non è stato delegato il compito di selezionare un preciso rispondente all'interno della famiglia¹¹¹; il suo ruolo è stato pressoché circoscritto alla consegna della busta.

Ai rilevatori è stato suggerito (non imposto) di svolgere l'attività di consegna in orari tali da aumentare la probabilità di trovare qualcuno in casa così da promuovere l'iniziativa presso la famiglia contattata. Immediatamente dopo il recapito della busta, il rilevatore doveva compilare la scheda di consegna (specifica per ogni busta/questionario) e riversare il più presto possibile (la sera stessa) le informazioni raccolte in un'identica scheda CAWI predisposta allo scopo. Più che un rilevatore lo si potrebbe definire un "postino intelligente". Il suo compito primario è stato

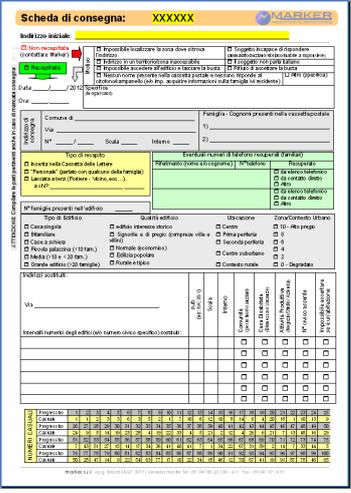
¹¹¹ Un solo componente della famiglia poteva rispondere al questionario. L'intervistato era, tra i familiari, colui che aveva la data di nascita più vicina ad una data casuale (del tipo giorno/mese) presente nella lettera di accompagnamento.

quello di cooperare attivamente nell'esecuzione dell'esperimento di campionamento (selezione della famiglia). In seconda battuta ha avuto il compito di promuovere la ricerca presso la famiglia selezionata e, nel contempo, raccogliere più elementi possibili su di essa e sul contesto di residenza. L'aspetto delicato di tale ruolo è che gli eventuali errori incidono sul carattere sperimentale della rilevazione.

Una prima e sostanziale contrazione dei costi è derivata dal contenimento dei tempi-lavoro dei "postini intelligenti". La somministrazione diretta del questionario (*face to face*), infatti, imporrebbe la presenza del rilevatore presso il nucleo familiare per una durata elevata (circa un'ora sommando anche i tempi interlocutori iniziali e finali). Inoltre, poiché la persona da intervistare è una precisa figura tra i componenti della famiglia, molto probabilmente sarebbero stati necessari più tentativi d'incontro. La scelta di adottare dei "postini intelligenti" non deriva solo della necessità di contenere i costi, ma anche dal tentativo di limitare l'influenza dell'intervistatore sulla rilevazione del dato, salvaguardando nel contempo la necessità di incrementare o far nascere una positiva empatia tra ricerca/rilevatore e intervistato. Favorendo la consegna diretta e personale del questionario, si è cercato di aumentare la credibilità della ricerca, cercando di far capire alla famiglia che, tra molte, lei e solo lei era stata selezionata in modo casuale e che la partecipazione era fondamentale allo scopo. Nessun altro avrebbe potuto esprimere esattamente la loro opinione su temi importanti quali la salute e l'ambiente. Un'occasione da non perdere! Superare la pigrizia generalizzata, se non anche la diffidenza della gente, è stato l'obiettivo principale dell'azione di contatto diretto i cui effetti saranno presentati in seguito (cfr. cap. 4.6). Un costo economico che si è reputato di dover sostenere. In alternativa, si sarebbe potuto impiegare il rilevatore per raccogliere semplicemente il nominativo della famiglia selezionata (e/o gli indirizzi validi in un tratto di strada), completare così l'azione di campionamento e spedire successivamente il materiale via posta ordinaria. L'operazione sarebbe stata meno onerosa e perfettamente in

linea con le esperienze internazionali di indagini postali del tipo *Address Based Sampling* precedentemente citate (Battaglia et al., 2008).

Figura 4.2: Attività del rilevatore e scopi da raggiungere

Attività del rilevatore	Scopi
<ul style="list-style-type: none"> • Individuare la famiglia pertinente a partire da un indirizzo fornito (o far decadere il punto di campionamento) • consegnare personalmente (se possibile) la busta-questionario e promuovere la ricerca; 	<ul style="list-style-type: none"> • Contrazione dei costi • Rafforzamento dell'empatia tra ricerca-rilevatore e famiglia intervistata;
 <ul style="list-style-type: none"> • recuperare i riferimenti telefonici della famiglia (presso le pagine bianche o direttamente) con cui compiere successivamente i solleciti alla compilazione del questionario; 	<ul style="list-style-type: none"> • Limitazione dell'influenza del rilevatore sulla varianza estranea al fenomeno indagato; • Rafforzamento/controllo dei meccanismi di casualità nell'individuazione della famiglia pertinente; • Predisposizione della fase di sollecito alla compilazione;
<ul style="list-style-type: none"> • compilare contestualmente la scheda di consegna cartacea; • riversare ogni sera i dati relativi alla scheda di consegna sul web 	<ul style="list-style-type: none"> • Controllo/certificazione e dell'operato del rilevatore (veloce verifica dell'andamento dell'indagine); • Acquisizione di dati contestuali all'indagine (relativi all'edificio e al contesto urbano);

Azione

Adeguate selezione e formazione della rete dei rilevatori

Ogni scheda di consegna è stata identificata da un codice alfanumerico cui corrispondeva una busta da recapitare. Nell'intestazione della scheda è stato riportato l'indirizzo (detto indirizzo iniziale) presso il quale il rilevatore ha dovuto recarsi. A seconda dell'esito della consegna (a buon

fine o meno, coerentemente ai motivi di mancata consegna), il rilevatore ha riportato le seguenti informazioni:

- data e ora dell'evento
- flag di conferma di avvenuta consegna
- motivo (modalità codificate) della mancata consegna e commento sull'accaduto
- indirizzo finale (in forma completa) presso cui si è stata tentata, o è avvenuta, la consegna¹¹²
- cognomi presenti nella cassetta postale
- modalità di consegna (diretta e personale, inserimento della busta nella cassetta postale, altro)
- riferimenti telefonici e modalità di reperimento dell'informazione (pagine bianche o ricevuti direttamente da un membro della famiglia)
- numero di nuclei familiari presenti nell'edificio
- tipo di edificio (casa singola, bifamiliare, ecc.)
- qualità dell'edificio (casa popolare, di pregio, tipica dell'area, ecc.)
- ubicazione dell'edificio (centro cittadino, prima o seconda periferia, località/frazione, ecc.)
- voto (da 0 a 10) sulla qualità del contesto urbano
- elenco degli edifici non eleggibili scartati prima di arrivare a quello finale presso cui la consegna è avvenuta o si è fatto decadere l'indirizzo
- Motivo di esclusione dell'edificio (o gruppi di indirizzi): comunità, casa disabitata, attività produttiva, ecc.

La qualità, la congruenza e la completezza dei dati inseriti nella scheda, o, meglio, nell'analogica scheda web/CAWI, sono stati puntualmente controllati, quasi per intero in modo automatico.

4.2.2 Caratteristiche desiderate per il rilevatore

Ogni potenziale candidato rilevatore doveva possedere alcuni requisiti di base irrinunciabili:

- risiedere nel comune oggetto d'indagine o conoscere molto bene il territorio¹¹³

¹¹² Vedremo più avanti quali sono i casi in cui indirizzo iniziale e di consegna non coincidono.

- non essere un professionista della ricerca, cioè non svolgere l'attività di rilevatore come professione primaria o secondaria (è stato comunque valutato positivamente il fatto di aver avuto qualche esperienza in tal senso)
- possedere almeno il diploma di scuola media superiore
- possedere i seguenti strumenti: pc, connessione internet, cellulare, automobile
- poter mettere a disposizione circa venti di ore di lavoro nell'arco di due o tre settimane in orari coerenti con i tipici momenti in cui le persone sono a casa (ore pasti e fine settimana).

4.2.3 Metodo e costo del reclutamento dei candidati rilevatori

La ricerca dei rilevatori è stata condotta tramite telefonate casuali tra le utenze dei comuni selezionati per l'indagine e attraverso internet consultando gli annunci di chi aveva pubblicato la propria disponibilità per attività di ripetizione. L'accreditamento dei rilevatori, effettuata prevalentemente tramite telefonate casuali, ha richiesto oltre 160 ore di lavoro concentrate nelle fasce serali (18:00-21:00). Per tale attività è stato d'ausilio un software CATI connesso ad un sistema CTI avente la funzionalità di *predictive dialing*, idoneo alla gestione efficiente degli appuntamenti e al controllo delle "quote" di candidati raggiunte in ogni comune. Sono state effettuate circa 7.400 chiamate usando 5.650 numeri. Sono state raggiunte 3.250 famiglie (circa il 57% dei numeri utilizzati) raccogliendo una disponibilità potenziale di oltre 120 soggetti in possesso dei requisiti minimali richiesti (circa il 3,7% delle famiglie contattate).

¹¹³ Tale richiesta nasce dalla convinzione che la consegna attraverso un concittadino sia accolta con minore reticenza (Boccuzzi, 1985) e che la conoscenza del territorio possa rappresentare un valore aggiunto nell'individuazione di possibili errori nel *database* geografico utilizzato per la selezione degli indirizzi.

Quantità che sale a 134 se si assommano anche gli esiti della ricerca in internet tramite annunci¹¹⁴.

4.2.4 L'accreditamento e la selezione delle candidature

Ai 134 potenziali rilevatori è stata inviata una mail chiedendo di:

- scaricare e leggere un documento con indicato in modo dettagliato l'impegno richiesto
- prendere visione delle norme sulla tutela della *privacy* per quanto riguarda il possesso da parte della società addetta alla gestione dell'indagine dei loro dati anagrafici e fiscali
- confermare la disponibilità all'incarico attraverso l'imputazione dei propri dati anagrafici e fiscali in una "maschera CAWI" predisposta allo scopo.

Una volta raccolte le disponibilità, ai candidati rilevatori sono state fornite, sempre via mail, le credenziali di accesso all'area operativa web, presso un portale dedicato, in cui era possibile:

- scaricare il Manuale per il rilevatore¹¹⁵
- "studiare" gli indirizzi di recapito dei questionari¹¹⁶
- visionare altri documenti messi a disposizione (il questionario; le "schede di consegna"; il contratto di lavoro, ecc.)
- effettuare delle prove d'inserimento dei dati acquisiti con le schede di consegna, ecc.

¹¹⁴ Alle 160 ore impiegate per la ricerca dei potenziali rilevatori tramite telefonate andrebbero aggiunte quelle per il reclutamento effettuato tramite annunci internet e quelle dedicate all'attività di supervisione, predisposizione e gestione della fase d'indagine.

¹¹⁵ Trattasi di un documento PDF di circa 30 facciate. Molti sono i temi trattati: il ruolo e il comportamento che il rilevatore deve assumere durante l'indagine; gli aspetti tecnico-operativi del lavoro; la descrizione e le istruzioni d'uso degli strumenti messi a disposizione; la spiegazione di come classificare gli edifici ...

¹¹⁶ Gli indirizzi sostitutivi non sono stati resi visibili ai rilevatori. Se per un preciso indirizzo la consegna falliva, allora solo quel singolo indirizzo sostitutivo veniva consegnato/reso visibile al rilevatore. In questo modo nessuno ha trovato conveniente "scartare" un punto di campionamento sapendo in anticipo il sostituto. L'indirizzo "ri-campionato", infatti, poteva essere in qualsiasi zona del territorio comunale.

Le attività preliminari svolte dal candidato rilevatore sono state pensate non solo quali momenti formali del processo di accreditamento, ma anche come test sulla familiarità nell'uso degli strumenti internet e come controllo del funzionamento dei mezzi informatici sviluppati allo scopo.

A causa dei ritardi nella partenza del progetto, alcuni candidati hanno dovuto declinare l'incarico. Altri, una volta valutato pienamente l'impegno richiesto attraverso lo studio del Manuale, hanno preferito rifiutare sottolineando che l'attività si presentava troppo onerosa rispetto alla remunerazione proposta e/o di aver sottovalutato l'impegno e la preparazione richiesti (l'accuratezza con la quale avrebbero dovuto svolgere il lavoro, i luoghi da presidiare, i tempi e lo studio preventivo del Manuale). La committenza universitaria e la tipologia dello studio (internazione e su tematiche come salute e ambiente) hanno favorito la credibilità dell'indagine e della società incaricata alla rilevazione e, dunque, la raccolta di candidature. In alcuni soggetti, addirittura, il nobile scopo ha ingenerato un atteggiamento più da volontario che da collaboratore-lavoratore. In linea generale ogni persona attribuisce un valore economico soggettivo al tempo che dedica a un'attività lavorativa. Nel calcolo entra a pieno titolo anche la libertà organizzativa concessa e/o percepita. È una conferma che il costo organizzativo assume un valore anche in ragione dei vincoli operativi che si adottano, riportandoci a quanto sviluppato nella prima parte della tesi dove è stato sottolineato l'effetto sulla qualità del dato indotto dal processo di lavoro (cfr. par. 1.2.3). Più sono stringenti e definiti i modi e i tempi dell'azione (come la libertà di scegliere l'individuo da intervistare, i tempi di evasione, le procedure da seguire prima di rendere eleggibile un indirizzo) più elevato è il costo indotto, ossia lo sforzo percepito dal rilevatore. Tradotto in termini economici, la platea di chi può essere interessato a svolgere l'attività aumenta quando si concede maggiore libertà d'azione o potrebbe essere più contenuta la ricompensa pattuita. Si prenda come esempio quanto succede per alcune indagini CAWI con gli *spare-panel*; gruppi di soggetti molto numerosi sono di volta in volta

coinvolti nell'indagine, ma non esiste vincolo alla partecipazione. Chi tra loro aderisce viene pagato e il tempo massimo entro il quale è possibile rispondere è dettato dal raggiungimento di un obiettivo numerico-campionario. La possibilità di partecipare all'indagine può durare poche ore come qualche giorno. Il meccanismo favorisce chi è disposto a "cogliere l'attimo"¹¹⁷. La flessibilità percepita porta a definire ricompense decisamente basse: in Italia si parla di pochi euro per ogni ora di collaborazione (tempo di auto-somministrazione del questionario). A volte, indagini con questionari brevi sono pagate più o meno un euro. Al contrario, il costo di un rilevatore CATI che intervista un cittadino, quest'ultimo ovviamente non remunerato, è gioco forza maggiore. Il CAWI con *spare-panel* ha quindi costi diretti di reperimento delle informazioni minori rispetto al CATI: la spesa è contenuta perché il trasferimento del costo intervistatore non è riversato in egual misura sul rispondente CAWI.

Tornando alla selezione delle candidature per ISSP Italia, degli oltre 130 potenziali rilevatori iniziali, si sono coperti i 60 punti campionari "scartando", perché in esubero, solo qualche persona (uno su due si è autoescluso dopo aver compreso le condizioni o, soprattutto, a seguito dello slittamento nella partenza dell'attività di rilevazione). L'attività di *screening* descritta in precedenza ha rappresentato un momento oneroso sia per il gruppo di ricerca sia per i potenziali rilevatori che hanno dovuto dedicare gratuitamente del tempo per valutare i documenti messi a disposizione. Il percorso di accreditamento ha costituito un deterrente per quanti erano attratti solo dal compenso e attribuivano poca importanza al

¹¹⁷ In un recente articolo riportante un sondaggio d'opinione presso la popolazione sui temi dell'immigrazione (febb. 2014), svolto nel veneto quale focus e in Italia quale alternativa di confronto, quasi in risposta a quanto appena affermato, si trova scritto circa la tecnica d'indagine e di campionamento: "metodo CAWI su campioni casuali probabilisti stratificati e di tipo panel ruotato". Con tutti i possibili errori del caso, tento di tradurre: "tra quanti collaborano sistematicamente alle nostre indagini (panel) rispondendo a questionari loro proposti via internet (CAWI), al fine di rendere simile alla struttura per genere e classi d'età della popolazione residente il gruppo dei rispondenti finali (campioni stratificati o per quote?), si è deciso di proporre l'intervista a dei soggetti scelti a caso presso le nostre liste (probabilistici casuali)." A mali estremi, estremi rimedi? Una sorta di "rotazione" imposta?

contenuto della prestazione: è stato un momento di valutazione della “buona volontà” e dedizione del candidato.

4.2.5 Alcuni aspetti legati alla formazione dei rilevatori

Per motivi strettamente economici non si sono predisposti web-test di verifica della preparazione dei rilevatori una volta studiato il Manuale operativo. Per ogni rilevatore, però, si è monitorato il *download* dei documenti e gli accessi alle strutture informatiche in visione (strumenti d’inserimento dati da utilizzare durante l’indagine). Se la verifica aveva esito negativo, si contattava la persona segnalando i passi necessari per l’auto-formazione. In sostanza, non si è verificata la comprensione dei contenuti del Manuale, bensì tutte le attività che il documento stesso indicava quali passaggi tecnici e opportunità di simulazione. La definizione di un web-test si configurerebbe comunque come una strategia migliorativa così come sarebbe opportuno affiancare (o sostituire) il Manuale a un video per ottimizzare i tempi di apprendimento.

Le domande su questioni specifiche poste dai rilevatori durante la loro formazione, e le relative risposte, sono state pubblicate nel portale (sezione FAQ) in modo che i dubbi di alcuni diventassero patrimonio conoscitivo comune. Per alcuni temi di rilievo si sono spedite mail specifiche a tutti poiché la sezione FAQ non discriminava l’importanza dei temi trattati e non era stata inizialmente concepita per gestire gli argomenti per aree tematiche. In ogni caso, il responsabile operativo del progetto poteva essere contattato in qualsiasi momento della giornata (dalle 9:00 alle 22:00) con una mail o una telefonata.

4.3 Il completamento del processo di campionamento

4.3.1 Eleggibilità ed ineleggibilità dell'indirizzo

Uno dei principali scopi del Manuale del rilevatore era quello di far comprendere il concetto di **eleggibilità o ineleggibilità di un indirizzo**, elemento strettamente collegato alla correttezza sperimentale del piano di campionamento. A partire da un indirizzo (via e numero civico), il rilevatore doveva:

❑ interrompere la consegna quando:

- in presenza di una residenza stabile (indirizzo eleggibile) si trovava di fronte ad una delle seguenti situazioni:
 - l'unico elemento della famiglia non era in grado di rispondere (problemi linguistici, malattia, anzianità, ecc.) o rifiutava la busta
 - risultava impossibile identificare la famiglia residente (nessun nome sulla cassetta delle lettere)
 - era impossibile accedere all'edificio o al territorio di riferimento dell'indirizzo (elevata criminalità, frane, ecc.)
- era impossibile localizzare l'indirizzo o l'indirizzo era riferito ad altro comune (scarsa qualità del *database* geografico)

Caduta della consegna

RICAMPIONAMENTO DELL'INDIRIZZO

❑ proseguire la ricerca di una residenza stabile (indirizzo eleggibile) quando, l'indirizzo in esame era ineleggibile, ovvero corrispondeva a:

- comunità (chiesa, albergo, ecc.)
- attività produttiva (verificando che non fosse presente una residenza per il custode o il proprietario)
- edificio disabitato (casa per vacanze, dismesso, in costruzione)
- numero civico non presente nella via.

Ricerca di un'abitazione stabile

Com'è facile intuire dalla casistica elencata, il recapito iniziale non garantiva sempre e comunque il reperimento di un nucleo familiare. Il civico fornito poteva non corrispondere a una abitazione, essere assente o, ancora, far riferimento a più nuclei familiari. Il rilevatore doveva operare attraverso regole rigide al fine di individuare, in prima battuta un edificio pertinente e, successivamente, nel caso di condomini, bifamiliari o complessi residenziali, selezionare un nucleo familiare tra quelli presenti.

4.3.2 Individuazione dell'edificio

Nel caso il numero civico fornito in prima istanza non fosse risultato eleggibile, il rilevatore doveva percorrere a ritroso la via rimanendo sullo stesso lato della strada. Ad esempio, se il civico iniziale era il numero 23, doveva verificare la presenza di un nucleo familiare (con residenza stabile) al civico 21, poi al 19 e così via. Se fosse arrivato alla fine della strada senza alcun esito positivo, doveva spostarsi sul lato opposto della via al numero 2 e da lì risalire (verificando i civici 4, 6, 8 ...) fino a trovare un punto di consegna. Nel caso tutta la via fosse risultata ineleggibile, doveva interrompere la ricerca, far decadere l'indirizzo e richiedere un altro punto di campionamento. Tutti i passaggi relativi agli edifici ineleggibili, inesistenti, ecc. dovevano essere riportati nella scheda di consegna, così come il motivo specifico della loro decadenza/sostituzione¹¹⁸. Tale procedimento si basa sul fatto che normalmente i lati destri piuttosto che quelli sinistri di una via hanno associata una numerazione pari e una dispari. Qualora ciò non si fosse verificato – la numerazione è continua come succede in una piazza piuttosto che in un vicolo –, il rilevatore doveva segnalare nella scheda il caso e arretrare seguendo la numerazione naturale (dal numero 5, ad esempio, passare al 4, al 3 e così via). Poco frequenti sono risultate anomalie del tipo:

¹¹⁸ È importante far notare che il termine sostituzione non deve essere inteso come scelta di un edificio eleggibile rispetto ad un altro egualmente eleggibile. Tale tipo di sostituzione è espressamente vietata. Il termine fa riferimento invece al processo di scelta, e conseguente eliminazione, degli edifici non pertinenti.

- in alcune piccole aree isolate, come i villaggi di montagna, i civici seguono la vetustà dell'edificio. Case con numerazione relativamente elevata si trovano dinnanzi o a fianco a edifici con numeri minori. Ogni nuovo edificio incrementa di una unità l'ultimo numero civico assegnato, indipendentemente dalla sua posizione all'interno del territorio o della via
- aziende agricole (o ex tali), situate lungo una carreggiata di collegamento tra due località, sono numerate in base alla distanza dell'edificio stesso dalla strada. Percorrendo la via si trova un viottolo che conduce ad un edificio dal civico del tutto imprevedibile (ad esempio, il 435). Esplorando il viottolo precedente o successivo, si può incontrare il civico 56 o 925. I numeri assegnati alle abitazioni senza un apparente criterio rappresentano i metri che separano l'edificio dalla strada principale
- la vecchia numerazione civica è ancora visibile e si somma a quella attribuita successivamente durante le periodiche revisioni effettuate dagli Uffici comunali. Analogamente, alcune abitazioni risultano prive di numero e inserite tra due civici consecutivi.

In tutti questi casi anomali, il rilevatore era tenuto a contattare il responsabile della ricerca il quale, molto spesso, sulla base della coordinata geografica generata per ottenere l'indirizzo dalla funzione di *reverse geocoding*, sceglieva la destinazione cui inviare il rilevatore. Le funzionalità di Google Maps e di *Google Street View* – la possibilità di “navigare fisicamente” entro i luoghi – rappresentano un ulteriore ausilio al coordinamento e alla verifica delle attività sul campo. Un sussidio tecnologico che pochi anni fa era inimmaginabile.

Un'ulteriore situazione trattata riguarda gli edifici identificati con un medesimo numero civico, ma distinti in base ad una lettera dell'alfabeto (es: 3/a; 3/b; 3/c). In questi casi la scelta dell'edificio prima, e del nucleo familiare poi, è avvenuta attraverso l'utilizzo delle tavole casuali.

4.3.3 Individuazione del nucleo familiare: l'uso delle tavole casuali

Ad ogni busta-questionario da recapitare, identificata da un preciso codice alfanumerico, era abbinata una scheda di consegna specifica (che riportava lo stesso codice alfanumerico) sulla quale era stampata una tavola di numeri casuali, differente per ogni punto di campionamento.

Figura 4.3: Facsimile di tavola dei numeri casuali presenti in ogni scheda di consegna

NUMERI CASUALI	Progressivo	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
	Casuale	1	1	2	3	3	6	3	5	2	1	7	10	6	12	10	8	14	6	4	20	18	1	10	13	9
	Progressivo	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
	Casuale	24	9	5	14	19	23	28	4	19	22	33	4	5	21	2	12	4	26	6	21	7	13	48	16	29
	Progressivo	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
	Casuale	7	43	51	27	15	14	17	34	36	41	40	41	7	1	34	22	67	31	59	2	14	7	33	44	5
	Progressivo	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
	Casuale	50	28	47	14	9	22	54	63	77	17	10	22	34	18	46	12	70	67	41	19	91	87	78	46	65

La tavola doveva essere utilizzata qualora fosse stato necessario operare una scelta tra diverse alternative: se l'edificio era caratterizzato da più ingressi (scala A, B, C); quando il complesso residenziale riportava lo stesso numero civico (3/a, 3/b); se l'edificio ospitava più nuclei familiari (condominio). In quest'ultimo caso, certamente il più diffuso, si doveva innanzitutto enumerare le alternative possibili (scartando ovviamente i locali aventi attività produttive come studi professionali e sedi di rappresentanza) e, successivamente, scegliere la famiglia in base alla tavola dei numeri casuali presente nella scheda. Era stata stabilita una convenzione per ordinare le alternative possibili, fornendo ai rilevatori numerosi esempi per limitare il più possibile scelte discrezionali.

4.4 Il fieldwork e i risultati ottenuti

4.4.1 La consegna del questionario

L'attività di consegna dei questionari alle famiglie è iniziata il 16 Ottobre 2012 e si è conclusa circa un mese e mezzo dopo (il 3 Dicembre). Ogni rilevatore doveva consegnare circa 65 buste¹¹⁹ ad altrettante famiglie all'interno del comune. È stato necessario ricampionare il 17% circa degli indirizzi iniziali. I motivi per i quali la consegna non è andata a buon fine sono riportati in Figura 4.4:

Figura 4.4: Distribuzione % delle sostituzioni per motivo

%		Motivo
39%	7%	Altro motivo (prevalenti le aree totalmente produttive)
28%	5%	Rifiuti
10%	2%	Cassetta delle lettere senza nome
8%	1%	Indirizzo sconosciuto – Area/territorio non localizzato
8%	1%	Edificio e/o area inaccessibile
6%	1%	Persona incapace di rispondere (quasi tutti anziani soli o con una badante presso l'abitazione)
1%	0%	Problemi linguistici
100 %	17%	Totale

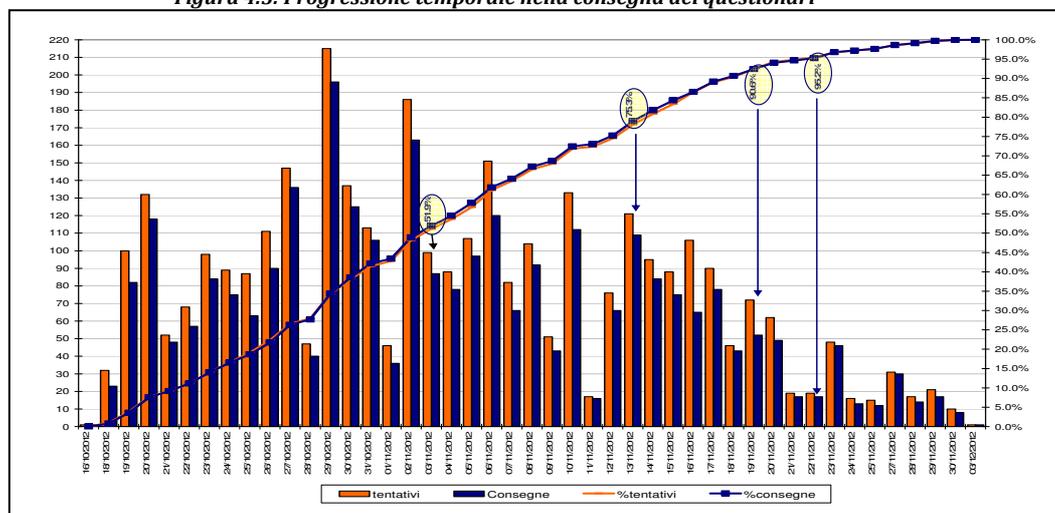
Entro i primi 18 giorni si erano già effettuate più della metà delle consegne (51,9%) e a distanza di un mese circa si è raggiunto il 90% delle consegne totali. La progressione delle consegne, come visibile nella Figura 4.5, è stata pressoché costante per il primo mese d'attività. La conclusione del lavoro, invece, si è protratta per ulteriori 15 giorni (dal 19/11 al 3/12) poiché qualche rilevatore ha iniziato e terminato il lavoro in ritardo.

Se si volesse calcolare un indicatore di “sofferenza organizzativa” in “indagini a distanza”, come è il nostro caso, tra la completa defezione di un rilevatore (e la sua conseguente sostituzione) e quanti non si sono dimostrati particolarmente attenti ai tempi richiesti (entro un mese dal ricevimento dei questionari), esso si attesterebbe attorno al 10-15% delle consegne totali. È importante parametrizzare l'operato del rilevatore (e il rispetto delle tempistiche ne è parte integrante) perché, come vedremo più

¹¹⁹ Ogni busta conteneva: il questionario, la lettera di presentazione, una busta preaffrancata per la restituzione del questionario compilato, un depliant di presentazione generale dell'iniziativa.

avanti risulta il principale elemento di variabilità in relazione al tasso di adesione all'indagine (cfr. cap. 4.6).

Figura 4.5: Progressione temporale nella consegna dei questionari



Fatte 100 le consegne valide:

- il 60,2% è avvenuto all'indirizzo iniziale¹²⁰; per i rimanenti casi è stato necessario provvedere alla sostituzione
- nel 46,5% dei casi c'è stato un contatto diretto con un familiare, per un altro 51,0% la busta è stata lasciata nella cassetta delle lettere e nel residuo 2,5% si è chiesto a parenti, portieri dello stabile o vicini di casa di consegnare la busta alla famiglia selezionata
- per il 46,3% delle famiglie si disponeva anche del numero telefonico recuperato direttamente o tramite le pagine bianche, valore che è salito al 58,1% una volta effettuata l'ulteriore ricerca sugli elenchi telefonici in fase di sollecito¹²¹.

I numeri telefonici recuperati si sono distribuiti in modo all'incirca uniforme tra coloro il cui indirizzo iniziale di consegna è stato anche quello

¹²⁰ A comporre il 60,2% partecipa anche quel 3,8% di indirizzi con piccole varianti rispetto all'indirizzo fornito (es: scala b, interno 23).

¹²¹ Le percentuali fanno riferimento al totale dei numeri a disposizione sul totale delle consegne. Includono sia i numeri telefonici corretti, ossia coerenti con la consegna effettuata, sia quelli errati (omonimie nel cognome della famiglia), sia quelli per i quali non si è potuto stabilire la loro correttezza (nessun sollecito effettuato o sempre liberi, occupati, ecc.).

finale (60,2% del totale) e le sostituzioni per ineleggibilità dell'indirizzo. Come prevedibile sin dall'inizio, reperire un numero telefonico è un evento indipendente dal processo campionario di eleggibilità dell'edificio. Ciò assume un importante significato dal punto di vista della pratica campionaria. Infatti, ricordando l'esperienza americana (Kumar et al. 2011) e, nello specifico, all'utilizzo del *web-mining* quale strumento di controllo a priori dell'eleggibilità dell'indirizzo (cfr. par. 3.4.1.2), risulta evidente che l'elenco telefonico non costituisce un ausilio sicuro per il campionamento: solo in meno della metà dei casi avrebbe fornito a priori un nominativo (l'intestatario dell'utenza telefonica) cui inviare il questionario per posta ordinaria. Inoltre, sarebbe rimasto un passaggio obbligato la selezione di uno dei nuclei familiari nei casi di condivisione del medesimo edificio (condominio). Questo conferma il livello di copertura medio offerto dall'elenco telefonico in Italia e pone un quesito su come il gruppo di ricerca americano abbia risolto tale questione: tutti gli indirizzi estratti per i quali non esisteva un numero telefonico sono stati rigettati? Come è avvenuta la selezione di uno tra i nuclei familiari residenti nello stesso edificio (se mai l'informazione presente nell'elenco telefonico fosse affidabile nel dare conto di ciò)?

4.4.2 Il controllo sistematico sull'attività del rilevatore

Al termine di ogni giornata di lavoro il rilevatore era tenuto a compilare via web le schede di consegna relative agli indirizzi "conclusi", sia per quelli con esito positivo (consegna effettuata) che per gli altri casi (mancata consegna). A seconda dell'intensità del lavoro svolto dai rilevatori, e comunque almeno una volta al giorno, è stata controllata, scheda per scheda, la correttezza del "percorso stradale" effettuato; per i casi in cui la consegna non era andata a buon fine veniva pubblicato sul sito il nuovo indirizzo "ricampionato". Tale serrato e sistematico controllo ha consentito ai rilevatori di chiarire eventuali incomprensioni relative alle regole studiate sul Manuale. Assai raramente ciò ha creato tensione tra le parti. Molto più spesso, al contrario, ha convinto il rilevatore che il proprio

impegno non fosse una pura formalità, ma un tratto distintivo della serietà e della qualità dell'indagine; aspetto che ha gratificato, spesso, l'impegno profuso.

La congruenza delle informazioni fornite dai rilevatori e la loro completezza, sono state verificate utilizzando una procedura informatica ad hoc a disposizione del supervisore. Il programma di controllo è stato concepito volutamente *offline*, separato dalla fase CAWI di inserimento dati¹²². Quando si è resa necessaria una ricampionatura dell'indirizzo o si è individuata qualche possibile incongruenza sui dati di consegna, è stata inviata una mail al rilevatore (scheda per scheda, cioè per ogni singola istanza da chiarire) il quale era quindi informato della presenza di un nuovo indirizzo da utilizzare (e delle operazioni da compiersi) o dell'argomento su cui era tenuto a confrontarsi con il supervisore (qualificare, correggere o chiarire). I dubbi sono stati spesso dissipati semplicemente con una telefonata tra le parti. Altre volte il rilevatore ha risposto alla mail e il supervisore, prendendo atto della correttezza della soluzione adottata, ha proceduto alla sistemazione della relativa scheda web. Si è spesso trattato di errate digitazioni, mancata segnalazione di edifici (sequenza di civici) nella scheda web, ecc. Al rilevatore non è stata offerta la possibilità di modificare in modo autonomo le schede web già inserite; gli interventi sui dati già imputati sono stati fatti dal supervisore dopo il confronto con il rilevatore stesso.

Questo tipo di organizzazione ha permesso di certificare la quantità delle consegne effettuate in palese violazione alle regole campionarie. Il tasso di errore finale è stato di poco superiore al 2% sul totale delle consegne. L'errore più diffuso è stato quello legato al concetto di casualità cioè attenersi ad una regola che individui casualmente un punto di partenza

¹²² L'idea nasce dalla convinzione di non ostacolare il libero inserimento dei dati. Se una maschera di imputazione pone dei vincoli di coerenza, ci potrebbe essere la tentazione di ricercare una soluzione indipendentemente dal fatto che questa corrisponda a ciò che veramente è avvenuto nella realtà. Si sarebbe ottenuta una formale, ma non sostanziale qualità del lavoro svolto. Inoltre, non si sarebbero evidenziati i tipici errori che i rilevatori tendono a commettere.

ed un rigido percorso stradale da seguire. Se un punto fisico vale l'altro, hanno pensato, perché non scegliere un edificio in base ad una "convenienza priva di malizia"? Così procedendo – "a casaccio" –, in alcuni casi si sono recati presso l'abitazione spazialmente più vicina rispetto a quella inizialmente loro comunicata (la casa accanto o di fronte). Il concetto di "casualità", figlio della rigida osservanza di una fase sperimentale di selezione, sembra essere il concetto meno compreso in assoluto, quello che tutti ritengono semanticamente simile al termine "a casaccio".

4.4.3 Il feedback verso il rilevatore

Nel Manuale è stato sottolineato quanto l'approccio diretto con le famiglie e la raccolta di numeri telefonici per un eventuale sollecito fossero due aspetti particolarmente importanti per il buon esito dell'indagine. Ad ogni rilevatore si è posto l'obiettivo di recuperare almeno il 50% dei recapiti (su fisso o cellulare) attraverso un contatto personale con la famiglia. La consegna personale era decisamente caldeggiata anche se non obbligatoria perché a volte presupponeva la necessità di fare più viaggi presso la medesima abitazione. Porre obiettivi stringenti in assenza di penali o congrui incentivi economici risulta decisamente difficile, sia perché i territori da presidiare sono differenti l'uno dall'altro (difficoltà di calcolo del reale sforzo profuso¹²³), sia perché il controllo sull'effettivo incontro tra rilevatore e famiglia può avvenire solo attraverso un ulteriore contatto (anche telefonico) con la famiglia coinvolta. Definire un'adeguata remunerazione in ragione della reale quantità di lavoro svolto è una tematica complessa e spesso non gradita dai rilevatori, soprattutto quando non si ha un rapporto di fiducia consolidato tra le parti. Nella fase di acquisizione delle candidature, d'altra parte, si era già rilevata la diffidenza suscitata in molti interlocutori. Essere contattati a casa da una "sconosciuta entità" di ricerca per assumere un incarico di lavoro non è cosa di tutti i

¹²³ È possibile applicare un incentivo a km, ma ciò non tiene conto dell'eventuale necessità di recarsi più volte allo stesso indirizzo e non è semplice trovare uno strumento che certifichi quest'ultimo aspetto.

giorni. Figuriamoci se si fosse anche preannunciato che parte della remunerazione sarebbe stata variabile e calcolata in base agli sforzi profusi e condizionata da altrettanti obiettivi da raggiungere. Questo è il motivo per cui il compenso è stato stabilito solo basandosi sul numero di questionari da consegnare e quindi uguale per tutti. La qualità del lavoro va in ogni caso incentivata e gli obiettivi richiesti devono essere resi credibili, cioè ritenuti plausibili con uno sforzo che ai più deve sembrare congruo rispetto al riconoscimento economico. Un confronto trasparente con l'operato degli altri colleghi rilevatori è sembrata, da questo punto di vista, la via migliore. Si è quindi predisposto l'invio sistematico di un report (file PDF) sui risultati che progressivamente si stavano ottenendo sia in termini di consegne sia relativamente al numero di questionari ritornati compilati. Inserendo nella comunicazione anche il livello di rispondenza delle famiglie, il rilevatore, che era impegnato solo nella fase di consegna delle buste, poteva verificare quanto il proprio lavoro fosse stato proficuo in termini di adesione delle famiglie all'indagine. Il report inviato ai rilevatori è stato predisposto secondo il seguente schema:

<ul style="list-style-type: none"> • Indirizzi totali forniti (iniziali + ricampionati) 	<ul style="list-style-type: none"> • n° di errori in violazione del metodo di campionamento 	<ul style="list-style-type: none"> • Numeri di telefono recuperati 	<ul style="list-style-type: none"> • Giorni medi trascorsi dalla consegna
<ul style="list-style-type: none"> • Consegne ultimate per tipo di recapito 	<ul style="list-style-type: none"> • Indirizzi "Scalati"¹²⁴ 	<ul style="list-style-type: none"> • Questionari restituiti completi o parziali (>33%) 	

Inoltre, sono stati calcolati (e spiegati in termini di significato) tre indici di correlazione tra i questionari compilati e:

1. il numero di consegne personali
2. la quantità di riferimenti telefonici recuperati
3. il tempo medio trascorso dalla consegna dei questionari alle famiglie.

¹²⁴ Il numero di indirizzi per i quali il rilevatore ha attivato la procedura campionaria di selezione dell'edificio eleggibile a fronte di un civico iniziale ineleggibile.

Relativamente al terzo indicatore, si specifica che, fino a quanto si mantiene su valori elevati (tendente all'unità), l'indagine doveva essere considerata ancora attiva (risultati non stabili) e che potrebbe anche non azzerarsi nel caso in cui l'operato dei rilevatori non sia omogeneo. In effetti, nel caso specifico alcuni hanno fornito il loro contributo in ritardo o hanno trascinato il proprio lavoro per un lasso di tempo più elevato. Elementi che generalmente sono sintomo e causa di un lavoro svolto in maniera più approssimativa; tanto che il coefficiente non si è completamente annullato. È ciò che precedentemente abbiamo definito "sofferenza organizzativa" e che ha coinvolto all'incirca un 10-15% delle consegne totali.

Il report e gli indici proposti, oltre ad evidenziare l'importanza di una consegna personale del questionario e la quota di adesione che si riesce a recuperare attraverso solleciti, elementi sui quali entreremo nel merito più avanti, assumono rilevanza divulgativa presso i rilevatori e, soprattutto, diventano un ulteriore strumento di controllo dell'attività sul campo. I dati inseriti nel report hanno permesso di verificare le differenze da comune a comune, ossia da rilevatore a rilevatore¹²⁵. Ad esempio, il fatto che tutti gli indirizzi iniziali forniti potessero andare a buon fine rappresentava una difformità rispetto all'andamento generale. Sono stati quindi monitorati soprattutto quei rilevatori che non hanno mai fatto "cadere" alcun indirizzo, così come coloro che hanno richiesto un numero elevato di sostituzioni. Volendo essere pienamente obiettivi, la quantità limitata di casi analizzati per singolo intervistatore (sessantacinque consegne) non ci ha permesso di configurare un indicatore solido; ciò nonostante le macroscopiche differenze di *performance* sono diventate ben presto visibili. Il fatto che il consolidamento dell'intero report può avvenire solo successivamente all'imputazione delle schede di consegna e data la velocità con la quale i rilevatori hanno potuto svolgere l'incarico (anche solo un paio

¹²⁵ La leggibilità del report è stata agevolata dall'utilizzo di colori che hanno messo in evidenza i dati di *performance* differenziali rispetto all'andamento complessivo. Si è rappresentata un'analisi dei profili mettendo in rilievo i risultati fortemente positivi, quelli sopra la norma, quelli da considerarsi sotto la norma e quelli fortemente negativi.

di giorni), non ci permette di definire tale pratica uno strumento operativo ottimale, capace di apportare sicuri e preliminari cambiamenti sull'operato dei rilevatori (*real time monitor*). Ci sembra di poter affermare che una adeguata formazione e un incentivo sul risultato rimane il metodo più affidabile per salvaguardare gli obiettivi da raggiungere. Un approccio alternativo e forse migliorativo potrebbe essere quello di imporre ai rilevatori di eseguire un numero limitato di consegne (dieci o venti al massimo), effettuare l'analisi anche in ragione dei risultati storici conseguiti e successivamente far proseguire il lavoro con gli aggiustamenti del caso. Situazione di fatto verificatasi con alcuni rilevatori e che ha permesso l'individuazione e l'eliminazione di alcuni errori per loro pressoché sistematici/frequenti. Non è stata però, come già detto, una condizione generalizzata e imposta.

4.4.4 I solleciti alla compilazione

Grazie al recupero dei numeri telefonici¹²⁶, a distanza di qualche giorno dal recapito del questionario è iniziata la fase di sollecito effettuata direttamente dall'istituto di ricerca incaricato dell'indagine. Tale operazione è stata svolta utilizzando un software CATI per la gestione dei contatti telefonici (schedulazione delle chiamate nel tempo, gestione degli appuntamenti e delle rubriche telefoniche, ecc.). Alla famiglia contattata si è chiesto innanzitutto se aveva ricevuto la busta/questionario; in caso negativo, si è verificato se l'indirizzo e il nome fossero coerenti con i dati anagrafici riportati dal rilevatore; in caso positivo, viceversa, si è chiesto se qualcuno all'interno della famiglia avesse preso visione del contenuto e si è cercato di ottenere un'indicazione di probabilità di partecipazione all'indagine. Qualora al soggetto non fosse risultato chiaro il motivo della chiamata o dell'indagine, si sono spiegati più nel dettaglio gli scopi della ricerca e le regole di selezione del componente familiare. In alcuni casi una

¹²⁶ I numeri telefonici sono stati raccolti dal rilevatore durante il contatto con la famiglia o consultando le pagine bianche. L'istituto di ricerca, successivamente, ha cercato di incrementarne la consistenza attraverso una ulteriore interrogazione delle pagine bianche.

stessa famiglia è stata contattata più volte. Il percorso si concludeva quando si riscontrava che il questionario era stato spedito o quando l'interlocutore manifestava l'intenzione di non essere più contattato. Nel corso della telefonata si chiesta anche la disponibilità a ricevere una mail che, oltre a ribadire gli argomenti trattati durante il colloquio telefonico, forniva un *link* ipertestuale per la compilazione CAWI del questionario. La modalità di compilazione via web, citata anche nei documenti recapitati, era sempre possibile attraverso la digitazione esatta dell'indirizzo internet e del proprio codice personale; tramite il *link* presente nella mail, invece, la procedura era semplificata perché l'instradamento verso il proprio questionario era automatico/diretto.

Figura 4.6: Esiti dei solleciti telefonici

Esito del contatto e relative affermazioni degli intervistati	Questionario Ricevuto				Totale
	Sì	No	Nessuna informazione	Mancati contatti	
Elimina Numero (fax, registro opposizioni, ecc.)-STOP				2.6%	2.6%
Nessuna Informazione - contatti rimasti in "appuntamento/liberi/ecc."			2.4%		2.4%
Q Non Ricev. - Indir. Errato (prob. err Tel) - STOP		5.8%			5.8%
Q Non Ricev. - Indir. esatto - Il tel -STOP		9.4%			9.4%
Q Non Ricev. - Indir. esatto - I tel - Rimasti in "appuntamento"		0.6%			0.6%
Q Ricev. - (Collabora) - Rimasti in "appuntamento"	1.8%				1.8%
Q Ricev. - (Forse Collabora) - Rimasti in "appuntamento"	0.3%				0.3%
Q Ricev. - (Forse Collabora) - Non Richiamare - STOP	2.4%				2.4%
Q Ricev. - Non collabora - STOP	14.6%				14.6%
Q Ricev. - nessuna informazione ulteriore - Nuovo Appuntamento	2.2%				2.2%
Q Ritornato (compilato, vuoto o parziale) - STOP	40.1%				40.1%
Q già Spedito (ma mai arrivato) - STOP	4.1%				4.1%
Q gettato (da rispedire)	0.5%				0.5%
RIFIUTO OSTILE - Nessuna Info - STOP			8.4%		8.4%
RIFIUTO OSTILE - Q Non ricev - Errori indirizzo-STOP		0.3%			0.3%
RIFIUTO OSTILE - Q Non ricev - STOP		2.0%			2.0%
RIFIUTO - Q Non ricev - Indi esatto - STOP		2.4%			2.4%
Totale	66.1%	20.4%	10.8%	2.6%	100.0%

Nella Figura 4.6 sono classificati i circa 2.000 casi per i quali si è tentato un contatto telefonico con la famiglia¹²⁷. Nel 20,4% dei casi i

¹²⁷ Il numero di telefonate complessive, relative ai 2.000 soggetti contattati, sono state quasi 6.000 (una media di 3 telefonate a famiglia). L'intera operazione ha assorbito circa 180 ore di lavoro-uomo.

rispondenti affermano di non aver ricevuto alcun questionario. In realtà, valutando il tipo di rifiuto (non si è avuto modo di colloquiare sufficientemente per acquisire le informazioni) e i casi in cui è assai probabile che il numero di telefono fosse relativo ad una famiglia estranea alla ricerca, il valore si dimezza. Analizzati questi episodi anche in relazione alla modalità con cui è stato consegnato il questionario, risulta verosimile che molte di queste affermazioni nascondano un velato rifiuto a partecipare all'indagine o, ancora più probabile, che le buste recapitate siano state gettate ancor prima di visionarne il contenuto. In sostanza, il lavoro svolto dai rilevatori è da ritenersi credibile. D'altra parte, sarebbe paradossale che il rilevatore fornisca un numero di telefono di una famiglia che successivamente dichiarerà di non avere ricevuto nulla. Potenziali comportamenti fraudolenti sarebbero più propriamente da ricercarsi in tutti quei casi in cui non si ha modo di contattare la famiglia registrata nella scheda di consegna.

In molti casi, contattare la famiglia ha significato prendere atto che il questionario era già stato compilato e/o spedito, a conferma del fatto che non per tutti è stato necessario un sollecito. Ciò, unito alla propensione riscontrata verso il cartaceo (dal momento che era già in loro possesso), ha portato ad un volume di mail inviate alquanto ridotto (4% del totale delle famiglie coinvolte, corrispondente all'8% circa delle famiglie contattate telefonicamente).

4.5 L'adesione all'indagine

4.5.1 Le caratteristiche delle residenze

In base ai dati raccolti nella scheda di consegna, si riscontra qualche lieve (ma non significativa) differenza del tasso di adesione rispetto alle tipologie abitative, nonché in relazione al contesto urbano.

Figura 4.7: L'adesione e i contesti abitativi (Valori percentuali)

	Risposto	Non Risposto	Totale
	100	100	100
Tipologia di abitazione (Pearson Chi-Square =0.045)			
Casa singola	48.6	44.1	45.4
Bifamiliare	12.3	12.8	12.7
Schiera	11.4	9.7	10.2
Meno di 10 fam	19.6	22.9	22.0
10-20 fam	5.8	8.1	7.5
Più di 20 famiglie	2.2	2.3	2.3
Tipo/qualità dell'edificio (Pearson Chi-Square =0.003)			
Di pregio	2.8	3.2	3.1
Normale	30.1	24.7	26.2
Popolare	54.1	61.2	59.2
Rurale / Tipico	9.2	8.4	8.6
Interesse Storico	3.8	2.4	2.8
Ubicazione dell'edificio (Pearson Chi-Square =0.04)			
C. suburbano (frazioni)	36.5	37.0	36.9
Centro cittadino	34.7	38.2	37.3
I° periferia	19.6	18.6	18.9
II° periferia	5.1	3.6	4.0
Contesto rurale	4.1	2.5	3.0
Voto Contesto Urbano (Anova Test =0.21)			
Media su scala 1-10	6.32	6.23	6.26

Nel loro complesso i risultati sono coerenti con le conoscenze storiche relative alla difficoltà di contatto in edifici medio-grandi dei grossi centri urbani e, aspetto non riportato in Figura 4.7, con i soggetti anziani e/o con bassi livelli di istruzione. Di ciò ne può essere indicazione indiretta il lieve differenziale relativo alla qualità (voto) del contesto urbano. L'immagine generale che se ne trae è una maggiore propensione alla collaborazione nelle aree cittadine di maggior pregio, probabilmente quelle di nuova urbanizzazione e poste ai margini dei centri urbani, nonché costituite per lo

più da case singole, bifamiliari e/o casette a schiera. Il tutto sembra indicare una famiglia giovane, con un tenore di vita medio/medio-alto, con un reddito coerente con livelli di istruzione più elevati rispetto la media della popolazione. Diventa poi cura del gruppo di ricerca, attraverso l'analisi dei questionari, confermare o smentire queste prime sensazioni provenienti dalle schede di consegna, ma ciò che rappresenta la questione centrale per un buon lavoro sul campo è comunque accertare che la distribuzione dei rispondenti non sia fortemente "distorta", ovvero concentrata oltremodo in particolari contesti abitativi. Questo, nell'indagine ISSP Italia non sembra essere accaduto.

4.5.2 Il ruolo del CAWI

Il **tasso di adesione è stato pari al 27,5%** sul totale delle buste consegnate. È da considerarsi un risultato in linea con le esperienze internazionali (Kaplowitz et al. 2004) e decisamente buono nel contesto italiano. Ricordiamo che l'indagine postale pilota condotta nel 2010 dall'Università di Alessandria, aveva registrato un tasso di ritorno pari al 17-18% circa (comprensivo dei questionari restituiti non compilati). Grazie all'articolato processo di controllo del lavoro sul campo e dei contatti successivi con le famiglie si è incrementata l'adesione di un 10% ottenendo un tasso di risposta di oltre 1,5 volte superiore al precedente.

Figura 4.8: L'adesione e i canali di risposta

Canale di risposta	Questionari Completi			Questionari con compilazione parziale (>33%)			Totale rispondenti	Totale Recapitato
	Posta tradizionale	web	Totale Complete	Posta tradizionale	web	Totale Parziali		
% sul gruppo di riferimento	84.0%	16.0%	100.0%	37.5%	62.5%	100.0%	27.5%	100.0%
% sul totale rispondenti	82.4%	15.6%	98.0%	0.7%	1.2%	2.0%	100.0%	
% sul totale recapitato	22.6%	4.3%	26.9%	0.2%	0.3%	0.5%	27.5%	100.0%

Come risulta evidente dai dati riportati in Figura 4.8, una larghissima parte dei questionari è pervenuta attraverso posta ordinaria (83% circa). Fatto 100 il numero dei questionari restituiti, solo un 16,8% circa ha

preferito usare il CAWI. Un risultato che in prima istanza sembrerebbe contennuto, ma considerando che la diffusione della banda larga in Italia copre il 40-50% delle famiglie, significa che circa un terzo del campione ha scelto il CAWI quale modalità di compilazione¹²⁸. Il web, almeno per ora, non costituisce un elemento essenziale in un'indagine diretta ma, probabilmente, la sua assenza potrebbe generare qualche nascosta distorsione dettata dalla mancata attenzione dei ricercatori nel favorire la più ampia libertà e comodità di risposta.

Concentrando l'attenzione sulle compilazioni parziali (comunque superiori a un terzo del questionario), si nota che il web consente di acquisirne in misura maggiore rispetto al cartaceo per il quale evidentemente è minore la tendenza a riconsegnare un questionario che, senza entrare nel merito delle motivazioni, non si è voluto completare. Con gli strumenti CAWI si recuperano quindi informazioni che generalmente andrebbero perdute.

Il web, nel suo complesso e non solo riferendosi al CAWI, rappresenta uno straordinario strumento per velocizzare la raccolta dei dati, utile per amplificare la credibilità dell'iniziativa di ricerca, nonché, come appena detto, strumento essenziale per recuperare parte delle informazioni che altrimenti andrebbero perdute. Non è però ancora uno strumento familiare ai più. È quasi paradossale pensare che l'Italia, paese europeo (se non addirittura a livello mondiale) nel quale si acquistano più *smartphone* (in termini relativi) è allo stesso tempo uno tra quelli che più a fatica usa internet come mezzo per l'interscambio informativo¹²⁹.

¹²⁸ Durante l'ultimo Censimento della Popolazione (2011) le risposte via web sono state il 33,4% (Istat, 2013).

¹²⁹ Una stima recente resa nota da AGCOM (Autorità Garante delle Comunicazioni e del mercato), indica che più di un terzo degli italiani non è mai entrato in internet (AGCOM, 2013).

4.5.3 Il *digital divide* e le geotipologie

Figura 4.9: L'adesione e i geotipi (Valori percentuali)

	Geotipi	Tasso di risposta	Di cui web	% web su risposte
1	Verso la provincia	27.3%	3.7%	13%
2	Vecchi poli industriali	32.6%	7.1%	22%
3	Profondo Sud	36.0%	3.0%	8%
4	Vecchia provincia *	14.0%	2.0%	14%
5	Disagio rurale	22.0%	2.0%	9%
6	Sussidi *	36.0%	0.0%	0%
7	Economia Alpina *	46.0%	8.0%	17%
8	Agrobusiness *	38.0%	16.0%	42%
9	Aree urbane	21.5%	4.7%	22%
10	Nuove aree industriali	31.3%	5.3%	17%
11	Cinture urbane	30.7%	6.0%	20%
12	Aree miste *	46.0%	0.0%	0%
13	Mare e monti (vacanze)*	82.0%	6.0%	7%
14	Sud urbano precario	23.6%	3.1%	13%
	Totale complessivo	27.5%	4.6%	17%

Uno dei punti fondamentali del campionamento proposto sono le geotipologie e la conseguente stratificazione adottata. Ciò non ha un impatto indifferente sull'ottimalità del campione: nel caso si verifichi un risultato trasversale tra gli strati la sua adozione è quantomeno superflua se non addirittura fonte di inefficienza campionaria. La capacità delle geotipologie nell'intercettare la varianza del fenomeno rappresenta perciò un ambito d'analisi, direi, dovuto. Dal punto di vista del *fieldwork* possiamo osservare i tassi di risposta e l'utilizzo del CAWI. Sebbene alcuni risultati (quelli relativi alle geotipologie contrassegnate da un * nella Figura 4.9) sono da attribuirsi anche al differente livello di impegno profuso dall'unico rilevatore utilizzato per la consegna, è altrettanto evidente che le componenti materiali e culturali del *digital divide* rappresentano ancora oggi un sostanziale ostacolo all'utilizzo del CAWI quale strumento unico per le rilevazioni. Si noti, ad esempio, il divario nella risposta via CAWI tra il gruppo costituito dai geotipi 9 e 2 rispetto al geotipo 14, tipologie tutte riferibili alle principali aree urbane del paese. Il primo gruppo, dove il tasso di risposta via web è stato più elevato, fa riferimento quasi esclusivamente al centro e nord Italia. Le aree del sud urbane (geotipo 14), simili al primo gruppo in termini dimensionali ma con un'economia più precaria, hanno risposto via web in percentuale decisamente più limitata. L'utilizzo del web

sembra essere (quasi) una prerogativa delle aree urbane e/o economicamente più solide (come l'“agrobusiness”).

Figura 4.10: l'adesione e le aree Nielsen (Valori percentuali)

	% risposta	di cui web	% web su risposte
Centro	21.3%	5.0%	23.4%
Isole	22.9%	3.1%	13.8%
Nord Est	26.2%	4.4%	16.7%
Nord Ovest	36.0%	6.0%	16.7%
Sud	25.3%	3.5%	13.8%
Totale complessivo	27.5%	4.6%	16.9%

Un ulteriore riscontro lo si può ricercare valutando i tassi di risposta via web per aree Nielsen. Il centro Italia, l'area che meno di tutte ha partecipato all'indagine, in proporzione ha utilizzato maggiormente il canale web. Questo riporta alla mente quanto successo più o meno vent'anni fa nel processo di diffusione della telefonia radiomobile. La disponibilità e l'adozione del telefono cellulare allora, e sembra ripetersi oggi per la banda larga, coinvolse fin da subito le aree centrali del paese per poi espandersi negli altri territori (nel nord d'Italia in prima istanza). In quest'ottica, le percentuali che ritroviamo nella terza colonna della Figura 4.10 (distribuzione delle risposte via web distinte tra Nord, Centro e Sud-Isole) sembrano quasi essere stabilite a tavolino piuttosto che delle evidenze empiriche. I dati ci dicono anche che la variabilità calcolata tra le aree Nielsen è minore rispetto a quella tra geotipologie. Ciò non è solo frutto di una statistica basata su di un numero minore di modalità, ma, molto più semplicemente, le aree Nielsen rappresentano degli aggregati territoriali (ripartizioni amministrative di livello superiore) meno efficienti nell'intercettare la variabilità del comportamento dei suoi abitanti rispetto alle geotipologie. Le geotipologie non sono costruite a partire da variabili di comportamento, di atteggiamento, di consumo o altro; al contrario, si basano sulle strutture socio-economiche delle unità amministrative di base (i comuni). È evidente che tali strutture agiscono da leve nel definire i comportamenti di consumo a livello medio/generale. È come un cerchio che si chiude: i territori acquisiscono le sembianze di quanti vi risiedono e le unità territoriali vanno a sostituire dal punto di vista analitico le unità

fisiche (i soggetti indagati) dando vigore alle analisi di geomarketing. Le geotipologie, in sostanza, rappresentano un piano interpretativo – paragonabile a ciò che viene svolto effettuando un'analisi delle corrispondenze – delle relazioni tra lo spazio, i fenomeni socio-economici e quelli di “consumo” (del web, in questo caso). È su tale capacità esplicativa che prendono forma le proposte di campionamento per ISSP Italia e per il caso americano (Kumar et al., 2011 e 2012). Se l'ottimalità del campione italiano potrebbe ancora essere messa in discussione dal punto di vista teorico, la scelta di operare in termini di geotipologie diventa cautelativa dal punto di vista della rappresentatività di un'Italia fortemente diversificata.

Un ultimo elemento che si vuole porre all'attenzione riguarda il caso dei comuni di cintura delle aree urbane dove il tasso di risposta è stato elevato, ma l'utilizzo del web si è stabilizzato su valori più contenuti rispetto alle aree urbane di riferimento. Ciò potrebbe essere una conferma della spinta autopropulsiva dei fenomeni socio-economici dettata dai processi di concentrazione, in questo specifico caso di tipo urbano (cfr. note 61-62-63).

4.6 Relazioni tra adesione, modalità di consegna e solleciti telefonici

Nell'ottica di migliorare le *performance* dell'indagine, è utile valutare l'impatto delle procedure poste in essere. Dato l'elevato costo sostenuto, è importante comprendere se la consegna personale tramite rilevatori e il sollecito telefonico rappresentano attività redditizia o se, diversamente da quanto ipotizzato, "il gioco non vale la candela". La risposta viene dall'indice di correlazione che è possibile determinare tra questi eventi. In particolare, l'indice di correlazione calcolato tra i tassi che fanno riferimento ad ogni rilevatore (numero di casi positivi su numero di consegne effettuate), propone i seguenti valori:

Figura 4.11: estratto del documento inviato al rilevatore (risultati definitivi conseguiti a distanza di qualche mese dall'ultima consegna svolta)

Indice di CORRELAZIONE tra il TASSO DI QUESTIONARI RITORNATI (completi e a compilazione parziale [>33%]) e il TASSO DI:

- Consegne Personali: 0.742
- Telefoni recuperati (totale): 0.656

Quando gli stessi indicatori sono calcolati sulle singole unità campionarie, la matrice di correlazione viene così valorizzata:

Figura 4.12: Matrice di correlazione tra ritorni, consegne personali e recupero dei numeri telefonici

	Consegne personali	Telefoni recuperati
Ritorni positivi	0,3175	0,3183
Consegne personali	-	0,2919

L'effetto positivo sul risultato dell'indagine è evidente e cospicuo così come è altrettanto chiaro che il contatto personale favorisce l'acquisizione del recapito telefonico. Per approfondire l'argomento, si può analizzare il tasso di adesione in ragione del tipo di approccio e di contatti tenuti con l'intervistato o la sua famiglia. Indicando con:

- *SollTel* sollecito telefonico
- *LinkByMail* sollecito via mail
- *TelFound* disponibilità di un recapito telefonico (comprensivo dei numeri errati)

si costruisce la seguente tabella:

Figura 4.13: Analisi delle modalità di consegna e impatto delle azioni di sollecito

TipoConsegna	TelFound	SollTel	LinkBymail	Esito positivo		
				% interviste Complete + Parziali (> 33%)		
Altro (2,5% del totale)	Vero	Vero	Vero	100%	15%(**)	24% (\$)
	Vero	Vero	Falso	13%		
	Vero	Falso	Falso	25%		
	Falso	Falso	Falso	(~) 6%		
Cassetta lettere (51,0% del totale)	Vero	Vero	Vero	75%	14% (**)	22% (\$\$)
	Vero	Vero	Falso	20%		
	Vero	Falso	Falso	25%		
	Falso	Falso	Falso	(~) 8%		
Personalmente (46,5% del totale)	Vero	Vero	Vero	78%	42% (*)	52% (***)
	Vero	Vero	Falso	49%		
	Vero	Falso	Falso	49%		
	Falso	Falso	Falso	(~) 18%		
Totale				27%		
	Vero	Vero	Vero	78%	39% (#)	15%
	Vero	Vero	Falso	36%		
	Vero	Falso	Falso	39%		
	Falso	Falso	Falso	11%		

Nota: i valori sono riferiti alle percentuali di riga a esclusione di quelli indicati nella prima colonna (TipoConsegna)

- **Effetto “contatto personale”** (differenziale di stima => Min. 15% - Max. 27-28%)

Avere un contatto diretto con la famiglia significa assicurarsi un ritorno medio del 42%(*). Negli altri casi (come recapitare il plico nella cassetta delle lettere) il tasso di risposta è simile a quello riscontrato nell'indagine postale di test (14 -15%(**)). Il contributo del contatto personale sul tasso di risposta varia tra il 15%¹³⁰ e il 27%-28%¹³¹. Che tale modalità di consegna faccia aumentare il livello di adesione all'indagine di oltre il 10% trova conferma anche dall'analisi dei tre sub gruppi “Falso_Falso_Falso”¹³² dove le quote di rispondenti passano dal 6-8%(~) (modalità “Altro” e “Cassetta lettere”) al 18%(~~) (modalità contattato personale). Il gruppo

¹³⁰ Valore ottenuto dalla differenza tra il 42% (tasso di risposta di chi è stato contattato personalmente) e il 27% (risultato generale).

¹³¹ Valore ottenuto sottraendo al 42% (tasso di risposta di chi è stato contattato personalmente) il 14-15% (tasso di risposta relativo alle altre modalità di recapito).

¹³² Falso_Falso_Falso, fa riferimento ai gruppi che dopo la consegna del questionario non hanno più avuto contatti nè con i rilevatori nè con la società incaricata al sollecito telefonico. Si tratta di gruppi numericamente poco consistenti.

18%^(~~) è verosimilmente composto da persone che, pur avendo un recapito telefonico (cellulare o fisso), non l'hanno voluto fornire o il loro atteggiamento era tale da scoraggiarne la richiesta. Anche su questo gruppo di "diffidenti", la risposta è maggiore di dieci punti percentuali rispetto ai similari "Falso_Falso_Falso" per i quali non è avvenuto un contatto personale.

Volendo monetizzare un contatto personale, esso dovrebbe essere retribuito da un 15% ad un 30% in più rispetto al caso in cui viene lasciato il plico nella cassetta delle lettere.

- **Effetto "sollecito telefonico"** (differenziale di stima => Min. 8-12% - Max. 24%)

Per quantificare tale effetto è opportuno prendere in esame il risultato totale (27%) e quel complessivo 39%^(#) relativo a quanti hanno ricevuto una chiamata (*SollTel* = Vero). Si tratta di un differenziale di circa 8-12 punti, pressoché analogo a quello che si osserva anche all'interno delle specifiche modalità di recapito. Possiamo notare che:

- per il recapito "Altro" la telefonata innalza il tasso di risposta dal 15%^(**) al 24%^(§)
- per chi ha ritirato la busta dalla cassetta delle lettere, il sollecito telefonico fa aumentare il tasso di successo dal 14%^(**) al 22%^(§§)
- per quanti hanno avuto una consegna personale e una successiva chiamata la percentuale di adesione passa dal 42%^(**) al 52%^(***).

L'effetto massimo del sollecito telefonico è circa 24% (differenziale tra chi è stato chiamato, 39%^(#), e il 15%^(###) mai contattato)¹³³.

L'analisi della Figura 4.13 ci fornisce almeno altre due indicazioni:

1. chi ha lasciato il proprio recapito mail, quindi presumibilmente interessato ad approfondire lo scopo dell'iniziativa e prendere informazioni circa i proponenti della stessa, ha in larga parte risposto (78% medio, gruppi "Vero_Vero_Vero"), indipendentemente dalla modalità iniziale di consegna del questionario. Purtroppo si tratta di un gruppo ristretto di casi (3-4% del campione).
2. per un interessante gruppo formato dai soggetti abbinati ad un numero telefonico, ma che non sono stati chiamati (tipologia "Vero_Falso_Falso"), la percentuale di successo risulta addirittura superiore a quella dei contattati telefonicamente. Questo perché il gruppo comprende quelli che hanno risposto immediatamente, magari via web, e comunque prima dell'avvio della fase di sollecito. Congiuntamente al gruppo precedente, questi rappresentano quelli che potremmo definire "entusiasti".

4.6.1 Azioni per aumentare l'adesione al progetto

Come aumentare il tasso di adesione? Rispondere a questa domanda e, nel contempo salvaguardare la qualità dell'indagine e contenere i costi, non è facile.

Un primo suggerimento potrebbe essere spedire delle lettere di sollecito (magari con allegato un'ulteriore copia del questionario) a coloro che non è stato possibile contattare telefonicamente. Questo gruppo nell'indagine ISSP Italia rappresenta una parte importante del campione,

¹³³ Gli intervalli (differenziali di stima), soprattutto se considerati nel loro punto massimo, sono da intendersi al lordo dell'effetto congiunto (correlazione) tra le due dimensioni analizzate ("sollecito telefonico" e "contatto personale") e al lordo del grado di disponibilità e fiducia della popolazione nei confronti di questo tipo di iniziative, dell'argomento trattato e delle istituzioni che la promuovono.

oltre il 42%. Tale attività era già stata sperimentata durante l'indagine postale di test con risultati discreti, ma non certo ottimali. Mettendo a confronto i tassi di risposta dell'indagine test (17% circa) con quelli qui ottenuti dal gruppo privo di chiamata telefonica (SollTel=Falso) e non contattato personalmente (8-9%) si può ipotizzare che una cospicua azione di solleciti postali potrebbe incrementare il tasso di adesione di circa 3-4 punti. Inoltre, spedire le lettere di sollecito consentirebbe di certificare gli indirizzi. Il servizio postale, infatti, è tenuto a restituire al mittente le lettere di indirizzi errati/sconosciuti. Quota parte della spesa potrebbe, di conseguenza, essere attribuita ai controlli di qualità sul *fieldwork* dell'indagine. Nel caso specifico si sarebbero dovute spedire oltre 1.600 lettere ed eventualmente i relativi questionari. Ad un costo minimo di circa 4-5 euro a busta (comprensivo del bollo e della stampa del questionario), significava una spesa ulteriore di circa 10.000 euro. Limitando alla sola lettera di sollecito, il costo sarebbe sceso a 2.000 euro. Probabilmente i 2.000 euro sarebbero un buon investimento, mentre i 10.000 euro sembrano francamente esagerati. Ciò sta ad indicare che perseguire tassi di successo sempre più elevati comporta una spesa crescente in modo non lineare (cfr. par. 1.3.1). Dal punto di vista organizzativo, il sollecito andrebbe spedito appena certificata dal rilevatore la consegna del plico/questionario. Considerando i tempi di recapito postali, infatti, il sollecito giungerebbe ad una distanza temporale ritenuta adeguata (5-6 giorni).

Un secondo suggerimento, forse quello che avrebbe maggiore successo, è quello di promuovere l'impegno profuso dai rilevatori con incentivi economici. L'aspetto è delicato perché si devono evitare meccanismi incentivanti che facciano aumentare l'ineleggibilità dell'indirizzo (abbandono dei casi difficili) e nello stesso tempo soppesare le diversità geotipologie (nelle grandi città spesso c'è una maggiore diffidenza verso gli estranei), nonché porre in essere dei meccanismi di controllo come sul numero reale di tentativi effettuati per realizzare una

consegna con contatto diretto con la famiglia. Non si tratta solo di presentarsi fisicamente dalla famiglia, ma anche di valutare il come si argomenta la ricerca e come si crea un'empatia positiva (Boccuzzi, 1985). "Venditori" non si nasce, ma nemmeno tutto si può pensare di far imparare facilmente il "mestiere" dell'intervistatore. Un approfondimento su come formare i rilevatori, nonché sugli strumenti di comunicazione messi loro a disposizione, è un passaggio doveroso.

4.6.2 Un patrimonio da valorizzare

La questione di come rendere più efficace l'azione di ricerca è lo scopo principale dell'analisi che si è sviluppata in queste pagine e che ha riguardato i dati e le pratiche adottate nell'indagine ISSP Italia. Lo si è fatto, ovviamente, assumendo che il piano d'azione rimanga sostanzialmente inalterato. Si è proposto di sperimentare l'adozione di una spedizione postale al fine di sollecitare le famiglie coinvolte per le quali un contatto telefonico è impossibile; ci si è convinti che sarebbe opportuno definire e concordare il metodo di remunerazione nei confronti dei rilevatori offrendo loro un sistema di incentivazione collegato al risultato finale; ed ancora, di come sarebbe necessario dare un riscontro mediatico all'iniziativa (ad esempio pubblicando i risultati della ricerca in internet).

Rimane invece ancora insondato il tema della valorizzazione di questo patrimonio nascosto (l'adesione avuta) creando, ad esempio, un gruppo sperimentale da non confondersi con i panel e con gli *spare-panel*. Le economie analitico-sperimentali da sviluppare potrebbero essere veramente importanti. Si potrebbero avere indicazioni circa le relazioni esistenti tra temi sondati con i vari moduli proposti da ISSP; si potrebbero analizzare le "cadute" di disponibilità che si registrano nel tempo in relazione ai diversi gruppi sociali. Si potrebbero avere, anche più semplicemente, maggiori opportunità per verificare quali pratiche operative consentono maggiori ritorni: telefonate o lettere di sollecito, questionari informatizzati più o meno interattivi, quantità di strumenti

(media) di risposta, ecc. Una sperimentazione tutta a favore dello sviluppo metodologico.

Conclusioni

Il termine tecnologia è spesso associato a quello di rivoluzione, quasi ad annunciare il fatto d'essere di fronte alla necessità di "buttare a mare" la vecchia cultura per far spazio a nuove idee, e nel caso specifico della ricerca sociale, a inediti strumenti analitici, metodologie e a nuovi approcci non probabilistici (Couper, 2013, "*Is the sky falling? New technology, changing media, and the future of Surveys*"). Ciò a partire dal fatto che, oggi più che mai, è cambiata la complessità, il volume e la natura dei dati con cui tentare di produrre una robusta informazione statistica. Non è forse questa, e per certi versi eccitante prospettiva, una prima conseguenza del mutamento tecnologico? La costruzione di una cultura scientifica non si concretizza in un esercizio censorio nei confronti di nuove idee e nuovi fatti. È, tuttavia e necessariamente, un percorso selettivo di governo del mutamento (de Jouvenel, 1964). In questo senso, se esiste una tendenza alla concentrazione dei dati nell'ambiente web, così da tentare stime sui risultati elettorali senza l'ausilio di un questionario, perché non sfruttare questa opportunità, nello specifico la rivoluzione geoinformativa (Kumar et al. 2011), per ampliare la dottrina del campionamento probabilistico anziché sottostare alla deriva metodologica connessa all'utilizzo degli *spare-panel* (rispondenti fidelizzati) a cui molti oramai sembrano essersi arresi? Questa sfida si è colta introducendo un campionamento probabilistico areale che sfrutta la funzione di *reverse geocoding* nell'indagine internazionale ISSP in Italia. Un metodo di campionamento "innovativo" che ha trovato uno sviluppo teorico, una sponda oltreoceano, esattamente nel periodo in cui l'ho progettato e applicato in Italia.

Parafrasando Schumpeter (1911) nel suo tentativo di dare un'anima alla dinamica economica, la tesi ha ulteriormente argomentato sul fatto che la tecnologia innesca un processo di "distruzione creativa"; trasforma cioè la realtà sociale attraverso nuovi strumenti e nuovi processi di produzione. È necessario prendere atto che essa incide anche sulle idee sociali e sulla

cultura scientifica. La trasformazione degli strumenti e delle tecniche di rilevazione sarà pressoché inevitabile, ma è importante prodigarsi affinché il mutamento produca positività e non ci si lasci condurre allo sbando (rifrazioni metodologiche). L'interazione esistente tra tecniche, strumenti di rilevazione e qualità dei dati è nota (Cocco e Tuzzi, 2012). Per molto tempo si sono prodotti studi a favore della standardizzazione dei processi di raccolta dati e la lunga storia del CATI, fatta di analisi *real time* dei tempi di somministrazione, di attenzione ai differenziali di stima e ai tassi di rifiuto per ogni rilevatore si inserisce proprio in questo filone culturale. Ora, a partire dalla "distruzione" della lista telefonica (strumento sul quale si basavano spesso i piani di campionamento probabilistici), sembra inevitabile il ricorso alle tecniche di rilevazione *mixed-mode*. I *mixed-mode*, quali processi produttivi, sempre più ricchi di multimedialità, tecnologia e varietà di forme, stanno seguendo un percorso di adattamento strumentale ai mezzi di comunicazione usati dalle persone. In questo cambiamento, ci si chiede spesso, però, che fine ha fatto la cultura della standardizzazione. È metodologicamente superata o, solo per motivi d'opportunità, destinata all'oblio? L'integrazione tecnologica degli strumenti di rilevazione è sorretta da una sensibilità tale da comporre un piano strategico e analitico finalizzato alla salvaguardia della qualità del dato rilevato? Nella tesi si è sostenuto come la ricerca applicata ponga spesso poca attenzione a come le opportunità tecnologiche possano essere sfruttate, ad esempio per tenere corti i legami tra ricercatori e rilevatori; una fonte di errore nota, a volte chiaramente denunciata (Boccuzzi, 1985), ma spesso trascurata. La proposta operativa del progetto ISSP Italia, con il suo impianto logico-organizzativo e con l'utilizzo di varie strumentazioni tecnologiche, ha cercato di coniugare economicità e velocità con la "vecchia" cultura di ricerca. Questo significa avere chiaro, innanzitutto, che la parcellizzazione dei processi produttivi non è sinonimo e neppure induce necessariamente all'industrializzazione. La tecnologia, nei processi di raccolta dati che stanno tentando di industrializzarsi per rendersi competitivi, dovrebbe

rappresentare un ausilio utile al contenimento dei costi transazionali, così spesso connessi alla salvaguardia della qualità e al buon “funzionamento” del prodotto finale (l’informazione). È un tassello importante nel dibattito metodologico in quanto la sola parcellizzazione porta invece ad espellere dalle “fabbriche dei dati” le professionalità “creative” dell’informazione, favorendo il sovrapporsi del concetto di dato con quello di informazione o, parafrasando il prof. Colombo, “il marmo con la scultura”. Senza una seria analisi dei modi con cui si possono materializzare i risparmi, parlare di competitività tra tecniche d’indagine e operatori del settore diventa un momento ricreativo nel mezzo di una impegnata giornata di studio. Parlare d’indagine “standard” o “rigorosa” (Keeter, 2000), affermare che esiste un “buon” modo di procedere in un’indagine CATI (Brusati 2001) e che il costo associato alle buone pratiche non differisce sostanzialmente da quello di altre tecniche di rilevazione percepite come costose, quindi “inefficienti”, è poco utile al ricercatore e alla cultura di ricerca se non si analizzano e si palesano i processi di lavoro connessi ai vari sistemi di produzione. Nella tesi un piccolo apporto a questo tema viene dato proponendo le curve di “costo metodologico”, ossia una valutazione della spesa connessa al raggiungimento di determinati tassi di adesione, un parametro riconosciuto da sempre quale indicatore di bontà di un’indagine. Un’idea semplice e riferibile ad un altrettanto comune sistema di valutazione che tutti noi spesso adottiamo: il “rapporto tra qualità e prezzo”. Questo approccio ci può aiutare nel confronto tra tecniche d’indagine, tra prassi di lavoro e nello studio dei tassi di adesione nel tempo. Inoltre, diventa un supporto per il ricercatore quando deve orientarsi nelle decisioni di spesa in funzione di ulteriori utilità marginali riferibili a un obiettivo di qualità auspicato.

Dare qualche risposta operativa, qualche spunto metodologico e fare una ricognizione concettuale sui processi di produzione del dato è ciò che si è cercato di fare in questa tesi. Si è messo in luce come la crisi del CATI altro non sia che la “distruzione”, indotta dall’avanzamento tecnologico, del

patrimonio di conoscenza sul quale basava la propria possibilità d'essere soprattutto strumento sperimentale: la lista delle utenze. In alcuni paesi, specie gli Stati Uniti, questi dati sono stati compensati dall'informatizzazione relativa alle consegne postali, i *Delivery Sequence File* (DSF). Il sistema di pressione dettato dai costi di ricerca ha trovato quindi uno spazio di compensazione nella sfera metodologica dando luogo a svariate iniziative di ricerca con campionamenti probabilistici e metodologie postali. In sostanza, la copertura dell'universo di riferimento è stata salvaguardata utilizzando una "creazione" indotta dal sistema tecnologico (i DSF). Qui l'eventuale *mixed-mode* viene mantenuto nel giusto ambito degli strumenti di raccolta dati e non è confuso, come spesso accade, con l'obiettivo strategico del campionamento. Il piano di campionamento applicato in ISSP Italia e quello proposto da Kumar (et al. 2011) per la General Social Survey del NORC nell'area metropolitana di Chicago, ossia l'introduzione della geoinformazione nella ricerca sociale empirica è da considerarsi un ulteriore avanzamento di questo modo di agire coerente con la dottrina classica della ricerca empirica. Lo spazio fisico diventa *layer* di interpretazione dei fenomeni sociali e, di conseguenza, anche utile strumento di predisposizione dei piani di campionamento. Nello specifico, il *reverse geocoding* assolve alla necessità di ricostruire l'universo di riferimento anche in assenza di una lista di indirizzi conosciuta a priori. Le indagini *mixed-mode* o semplicemente postali, possono in questo modo godere di un piano di campionamento probabilistico al pari di quanto accade con i DSF e di quel che accadeva con il CATI quanto la diffusione telefonica era pressoché totale. Un passo avanti fondamentale, sicuramente perfettibile, che copre una limitata parte delle necessità della ricerca empirica, ma meritevole di ulteriori applicazioni e approfondimenti.

Riferimenti Bibliografici

- AGCOM (2013), *Relazione annuale 2013 sull'attività svolta e sui programmi di lavoro* in www.agcom.it
- Banks R., Currall J., Francio J., Gerrard L., Khan R., Macer T., Rigg M., Ross E., Taylor S., Westlake A. (2003), "The Impact of Technology on the Survey Process: Proceedings of the Fourth ASC International Conference", *Association For Survey Computing*, Chesham
- Battaglia M. P., Link M. W., Frankel M. R., Osborn L., Mokdad A. H. (2008), "A comparison of Address-Based Samplig (ABS) versus Random-Digit Dialing (RDD) for General Population Survey" in *Public Opinion Quarterly* (72-1)
- Bernardi L. E. (2001), "Statistica e l'immaginario collettivo" in Tuzzi A. (a cura di), *Dall'intervista alla notizia*, Padova, Ed. Sapere
- Bernardi L. E. (2001), "Statistica e mezzi di comunicazione di massa" in Tuzzi A. (a cura di), *Dall'intervista alla notizia*, Padova, Ed. Sapere
- Bernardi L. (2005), *Percorsi di Ricerca Sociale*, Carocci Editore, Roma
- Bernardi L., Tuzzi A. (2005), "Questionario e question wording." in Bernardi L. (ed) *Percorsi di Ricerca Sociale*, Carrocci Editore, Roma
- Bianco M. L. (2009), "L'International Social Survey Programme (ISSP) e la ricerca comparativa nel tempo e tra nazioni", BANDO PRIN 2009, mod. A, MIUR – *Ministero dell'Istruzione dell'Università e della Ricerca*
- Bishop G.F., Hippler H.J., Schwarz N., Strack F. (1988) "A Comparison of response effects in self-administered and telephone surveys." in *Telephone Survey Methodology*, Wiley, New York
- Boccuzzi E. (1985), "Il punto di vista di un intervistatore" in Tuzzi A. (a cura di), *Dall'intervista alla notizia*, Padova, Ed. Sapere – Titolo originario "Parcellizzazione e reificazioni nelle ricerche sociologiche: il punto di vista di un'intervistatrice", *Rassegna Italiana di sociologia*, a, XXVI, n.2 aprili-giugno 1985
- Booth J. G., Presnell B. (1999) "Resampling Methods for Sample Surveys" in *Bulletin of the International Statistical Institute* (Finland), <http://www.stat.fi/isi99/proceedings/arkisto/varasto/boot0244.pdf>
- Bordoni C. (a cura di) (2013), *Communitas. Uguali e diversi nella società liquida*, Alberti Editore, Reggio Emilia
- Brusati E. (2001), "I campioni per i sondaggi demoscopici" in Tuzzi A. (a cura di), *Dall'intervista alla notizia*, Padova, Ed. Sapere – Titolo originario "La formazione dei campioni per i sondaggi demoscopici della Doxa" in Fabbris L. (a cura di)(1990), *Rilevazioni per campione delle opinioni degli italiani*, Padova, SGE Editoriali
- Campostrini S. (1990), "Ruolo ed effetti dell'intervistatore CATI" in Fabbris L. (a cura di), *Rilevazioni per campione delle opinioni degli italiani*, Padova, SGE Editoriali

- Cocco M. e Tuzzi A. (2012), "New data collection modes for surveys: a comparative analysis of the influence of survey mode on question-wording effects" in *Quality & Quantity*, October 2013, (47-6)
- Couper M. P. (2000), "Web surveys: a review of issues and approaches" in *Public Opinion Quarterly*, (64-4)
- Couper M. P. (2008), "Designing Effective Web Surveys" in *Cambridge University Press*, Cambridge
- Couper M.P. (2013), "Is the Sky Falling? New Technology, Changing Media, and the Future of Surveys" in *Survey Research Methods*, (7-3)
- Crescenzi F. (2002), "I Sistemi Informativi Geografici e la qualità del dato statistico" per *Sesta conferenza nazionale di statistica*, Roma
- de Jouvenel B. (1964), *L'arte della congettura*, Vallecchi (1967) Firenze, pp. 325-335
- De Mucci R. (1989) *La forma scienza. Scienza e scienze sociali*, Cleup , Padova.
- Di Gioia R. (2009), "Nel nome del campione rappresentativo: il CATI e il vitello d'oro", in *Sociologia e Ricerca Sociale*, n. 90, pp. 99-123.
- Diana G., Salvan A. (1993), *Campinamento da popolazioni finite*, Cleup , Padova.
- Dillman D. A., Phelps G., Tortora R., Swift K, Kohrell J., Berck J., Messer B.L. (2009) "Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet" in *Social Science Research* (38-1)
- Dillman D.A., Clark J.R., West K.K. (1995), "Influence of an invitation to answer by telephone on response to census questionnaires" in *Public Opinion Quarterly* (58-4)
- Eckman S. (2012), "*Uses of GIS in large-scale Surveys. Measuring People in Place*", Boulder CO, IAB (Institute for Employment Research – Federal Employment Agency)
- Eckman S., English N. (2012a), "Geocoding to create Survey Frame" in *Survey Practice* (5-4).
- Eckman S., English N. (2012b), "Creating Housing Unit Frames from Address Databases, Geocoding Precision and Net Coverage Rates" in *Field Methods* (24-4)
- Fabbris L. (1983) *Analisi esplorativa dei dati multidimensionali*, CLEUP, Padova
- Fabbris L. (1990), *Rilevazioni per campione delle opinioni degli italiani*, Padova, SGE Editoriali
- Fuchs M., Busse B. (2009), "The Coverage Bias of Mobile Web Surveys Across European Countries" in *International Journal of Internet Science* (4-1)
- Galbraith J. K. (1958), *The Affluent Society*, Houghton Mifflin Company, New York
- Gauthier A. H. (2000), "The Promises of Comparative Research" in *Paper for European Panel Analysis Group*.

- Gili G. (2005) *La credibilità. Quando e perché la comunicazione ha successo*, CZ, Rubbettino Editore, pp. 3-4
- Groves R. M., R. L. Kahn (1979), *Surveys by telephone: A national comparison with personal interviews*, New York: Academic
- Harkness J. A. (2003). "Questionnaire Translation" in *Cross-Cultural Survey Methods*, (a cura di) Harkness, J. A., F. J. R. van de Vijver, P. Ph. Mohler, Hoboken NJ: Wiley
- Hirschman A. O. (1990), *Come far passare le riforme*, BO, Il Mulino, pag. 281
- Iannacchione V. G., McMichael J. P., Shook-Sa B. E., Morton K. B. (2012), *A proposed hybrid sampling frame for the national survey on drug use and health. Final Report*, (RTI International - North Carolina)
- Iannacchione V. G., Ridenhour J.L., McMichael J. P., Shook-Sa B. E. (2010a), "The implications of geocoding error on address-bases Sampling" (RTI International - North Carolina) in *Survey Research Methods*
- Iannacchione V. G., Ridenhour J.L., McMichael J. P., Shook-Sa B. E. (2010b), "Linking Mailing Address to a household Sampling Frame based on Census Geography" (RTI International - North Carolina) in *Survey Research Methods*, pp. 3303-3311
- ISTAT (2006), *La rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione*, Metodi e Norme n. 32, Roma
- ISTAT (2013), *Annuario Statistico Italiano 2013*
- ISTAT (anni vari), *Indagine multiscopo "Aspetti della vita quotidiana"*
- Kaldor N. (1975, prima edizione 1972), "Irrilevanza della teoria dell'equilibrio economico" in F. CAFFÈ (a cura di), *Autocritica dell'economista*, BA, Laterza
- Kaplowitz M. D., Hadlock T. D., R. levine R., (2004), "A Comparison of Web and Mail Survey Response Rates", in *Public Opinion Quarterly* (68-1)
- Kasprzyk D., Duncan G., Kalton G., Singh M. P. (1989), *Panel surveys*, Wiley & Sons, New York -Chichester
- Keeter S., Kennedy C., Dimock M., Best J., Craighill P. (2006), "Gauging the Impact of Growing Nonresponse on Estimates from a National RDD Telephone Survey" in *Public Opinion Quarterly* (70-5)
- Keeter S., Miller C., Kohut A., Groves R.M., Presser S. (2000) "Consequences of Reducing Nonresponse in a National Telephone Survey" in *Public Opinion Quarterly* (64-2)
- Kim J., Gershenson C., Glaser P., Smith W. (2011) "The Polls—Trends In Surveys On Surveys" in *Public Opinion Quarterly* (75-1)
- Kreuter F., Presser S., Tourangeau R. (2008), "Social desirability Bias in CATI, IVR, and web surveys: the effects of mode and question sensitivity" in *Public Opinion Quarterly* (72-5)

- Kumar N., Jin Chen J., Liang D. (2011), *“Reverse Geocoding – a New Method of Enumeration Listing”*, Department of Geography, University of Iowa
- Kumar N., Liang D., Linderman M. (2012), *“An Optimal Spatial Sampling Design for Social Surveys”*, Division of Environment and Public Health, University of Miami
- Lee U. (2003), *“Panel Attrition in survey data: a literature review”*, working paper n. 41 Centre for Social Science Research, University of Cape Town
- Neyman J. (1938), *“Contribution to the theory of sampling human population”* in *Jour. Amer. Assoc.*
- Rao A. (2008), *“Come uscire dalla depressione sociale. Saggio sul pensiero economico e il mutamento sociale”*, Cleup SC, Padova
- Ridenhour J.L., McMichael J. P., Shook-Sa B. E. (2008), *“A robust procedure to supplement the coverage of address-based sampling frames for household surveys.”*, (RTI International - North Carolina) in *Survey Research Methods*
- Roccatò M. (2003), *Desiderabilità Sociale e Acquiescenza. Alcune Trappole delle Inchieste e dei Sondaggi*, LED Edizioni Universitarie
- Saris W. E., Gallhofer I.N. (2007) *Design, Evaluation, and Analysis of Questionnaires for Survey Research*, Hoboken, Wiley
- Sartori L. (2006). *Il Divario digitale: Internet e le nuove disuguaglianze sociali*, Il Mulino, BO.
- Schumpeter J. A. (1932, prima edizione tedesca 1911), *“La teoria dello sviluppo economico”*, in *Dinamica Economia, Vol V*, della Nuova Collana di Economisti, To, UTET
- Schumpeter J. A. (1962, prima edizione tedesca 1911), *“Scienza e ideologia”*, in F. CAFFÈ (a cura di), *Economisti moderni*, MI, Garzanti
- Schumpeter J. A. (1977, prima edizione tedesca 1939), *Il processo capitalistico, cicli economici*, To, Bordigheri
- Sraffa P. (1937, prima edizione 1928), *“Le leggi della produttività in regime di concorrenza”*, in *Economia pura*, vol. IV della Nuova Collana di Economicisti, To, UTET
- Stallman R. (2008), *“Cloud computing is a trap”*, intervista rilasciata al The Guardian, London
- Tuzzi A. (2001), *Dall'intervista alla notizia*, Padova, Ed. Sapere
- VandeVen A. H and Delbecq A. L., (1974). *“The Effectiveness of Nominal, Delphi, and Interacting Group Decision Making Processes”*, *The Academy of Management Journal* (17- 4)
- Weber M, Denk M, Oberecker K, Strauss C, Stummer C. (2008) *“Panel surveys go mobile”* in *International Journal of Mobile Communications*, vol. 6.
- Williamson O.E. (1987). *Le istituzioni economiche del capitalismo. Imprese, mercati, rapporti contrattuali*, Franco Angeli, Milano

Young A. (1928), "Increasing Returns and Economic Progress" in *"The Economic Journal"*

Siti consultati (*ultima data di consultazione: 19 Agosto 2014*)

http://en.wikipedia.org/wiki/Interactive_voice_response

http://en.wikipedia.org/wiki/Voice_stress_analysis

http://it.wikipedia.org/wiki/Big_data

voicesfromtheblogs.com/2013/02/26/elezioni2013-tra-sorprese-e-debacle-twitter-poll-meglio-dei-sondaggi/

<http://www.blaise.com/doc/doc5/>

<http://www.garanteprivacy.it/> (provvedimenti vari)

<http://www.gesis.org/en/issp/search-and-data-access/>

<http://www.issp.org>

<http://www.Lie-Detection.com> (X13-VSA Ltd)

<http://www.limesurvey.org/>

<http://www.ornl.gov/sci/landscan>

<http://www.registrodelleopposizioni.it/> (provvedimenti vari)

<http://www.rti.org/brochures/cari.pdf>

<http://www.youtrend.it/2013/01/27/twitter-intenzioni-voto-metodologia/>

<https://www.google.com/maps/>

Ringraziamenti

A chiusura di questo lavoro voglio ringraziare alcune persone per la collaborazione che mi hanno offerto e per la disponibilità dimostrata. Inizio citando la professoressa Cinzia Meraviglia per l'ottimo clima di lavoro che ha saputo creare negli oltre due anni nei quali abbiamo collaborato all'iniziativa di ricerca ISSP Italia. A Lei un grazie sincero per avermi affiancato nelle scelte, sempre incoraggiato ad affinare, innovare e, soprattutto, per aver riposto fiducia nella mia proposta di campionamento e di attuazione dell'indagine. Un esperimento metodologico il cui esito non era per nulla scontato.

Ringrazio Delfina Di Monte per tutte le volte che mi ha incitato a terminare il mio percorso di studi e per quanto ha saputo stupirmi per cortesia e disponibilità nelle mie oramai rare visite al Dipartimento di Statistica.

Ringrazio la mia relatrice professoressa Arjuna Tuzzi, per aver messo a disposizione la sua competenza, il suo tempo, pazienza e professionalità, ma soprattutto per aver accettato di inserirsi in un lavoro di tesi già nella sua fase conclusiva.

Voglio infine ricordare con profonda stima e grande affetto il professor Lorenzo Bernardi. Ho avuto la fortuna di conoscerlo come docente. Mi ha dato l'occasione di confrontarmi con altri suoi studenti intervenendo a qualche sua lezione quando uno dei temi interessanti nella ricerca sociale era il metodo di rilevazione CATI. Ho attinto alle sue competenze nell'ambiente di lavoro e, non ultimo, ho avuto modo di abbozzare con lui questo lavoro di tesi. Ciò che ho sempre ammirato in lui è il profondo rispetto e curiosità intellettuale che concedeva ad ogni suo interlocutore, me compreso.

Ci sono tante persone che hanno arricchito e che arricchiscono la vita di ognuno di noi. Spero di aver saputo e di avere modo di ricambiare la loro affettuosa presenza. Purtroppo con il professor Bernardi non ho avuto occasione. Per questo a lui dedico la mia tesi.