

Il trattamento dei termini composti nei thesauri

Luisa Venuti

Biblioteca "Mario Rostoni"
Università C. Cattaneo-LIUC, Castellanza
luisa.venuti@gmail.com

Scomponibilità dei sintagmi in base agli standard ISO e BS

1. Premessa

In questo articolo cercheremo di spiegare in che modo i due standard ISO 2788:1986 e BS 8723-1/2:2005, punto di riferimento per l'attività della Biblioteca della LIUC di costruzione del Thesaurus di economia e scienze sociali (ThESS),¹ affrontano una delle operazioni più difficili e delicate nell'ambito dell'indicizzazione semantica: il trattamento dei termini composti.²

Illustreremo i criteri in base ai quali un sintagma è considerato scomponibile, mettendo in luce le ambiguità che emergono dal testo dei due standard, dovute da un lato all'impiego con un significato non univoco dei termini relativi agli oggetti (termini e/o concetti) su cui si esercita l'attività degli indicizzatori e di chi costruisce thesauri, dall'altro, più a monte, all'assenza di una riflessione teorica coerente su ciò che i thesauri effettivamente organizzano, come mostra per esempio il fatto che entrambi gli standard abbiano assunto la teoria della significazione dei concetti da parte dei termini senza risolverne o, almeno, discuterne gli aspetti più problematici.

Alla base della nostra interpretazione degli standard vi è la convinzione che il thesaurus, come ogni vocabolario, non organizzi semplicemente i termini in base al loro significato, ma organizzando inevitabilmente *produca* significato, per cui fare thesauri è di fatto fare terminologia. La pratica della catalogazione semantica che si appoggia all'uso dei thesauri, o dei soggettari, mostra come la scelta e la collocazione di termini nelle gerarchie e nelle reti di relazioni

all'interno dei thesauri sia un'operazione culturale, che, attingendo al patrimonio terminologico offerto nei documenti afferenti a specifiche discipline di studio, e all'uso linguistico comune, con i limiti imposti dalle specifiche esigenze di ogni agenzia catalografica, crea uno spazio semantico all'interno del quale il catalogatore è portato a muoversi per elaborare il soggetto dai documenti.

Note terminologiche

Per chiarezza d'esposizione, nel corso di questo articolo con "termini" ci riferiremo selettivamente ai termini di indicizzazione, cioè ai termini³ controllati per uso bibliografico; con termini composti, senza richiamare in alcun modo il processo di composizione descritto dalla linguistica, indicheremo sintagmi in cui si possono distinguere diverse parti componenti, ciascuna significativa in sé ai fini dell'accrescimento del thesaurus, estratti dai documenti ai fini dell'indicizzazione;⁴ per evitare ambiguità impiegheremo "sintagmi" quando vorremo riferirci in generale alle unità della struttura sintattica degli enunciati; infine, "concetti" sarà utilizzato nel significato definito dal British Standard (v. sotto, par. 2.1). Termini prodotti a scopo esemplificativo saranno espressi tra virgolette, al plurale e con lettera iniziale maiuscola (es. "Autostrade"); concetti prodotti a scopo esemplificativo saranno espressi facendo ricorso a sintagmi posti tra virgolette, al singolare e con lettera iniziale minuscola (es. "autostrada" = "strada per la circolazione delle automobili").

¹ I due standard sono stati analizzati nell'edizione in lingua inglese: INTERNATIONAL STANDARD ORGANIZATION, *International standard ISO 2788: documentation: guidelines for the establishment and development of monolingual thesauri*. 2nd ed. Geneva, International organization for standardization, 1986; BRITISH STANDARDS INSTITUTION, *Structured vocabularies for information retrieval: guide. Part 1 (BS 8723-1): Definitions, symbols and abbreviations*, London: British standards institution, 2005; BRITISH STANDARDS INSTITUTION, *Structured vocabularies for information retrieval: guide. Part 2 (BS 8723-2): Thesauri*, London: British standards institution, 2005. Limitatamente ad alcuni casi specifici, il testo in lingua inglese è stato confrontato con la traduzione in lingua italiana: ENTE NAZIONALE ITALIANO DI UNIFICAZIONE, *UNI/ISO 2788: documentazione: linee guida per la costruzione e lo sviluppo di thesauri monolingue*. Milano: UNI, 1993.

² Cfr. ISO 2788:1986 § 7.1.1: "The establishment of procedures for dealing consistently with compound terms introduces one of the most difficult areas in the field of subject indexing"; viceversa, cfr. BS 8723-2:2005 § 7.3.2: "The decision on admission [Ndr, of complex concepts] is often difficult and subjective. The thesaurus developer should consider the balance of advantages in the retrieval situations that are likely to occur".

³ Per "termini" in generale intendiamo sintagmi "terminali", cioè ritenuti non ulteriormente scomponibili senza che si verifichi una perdita di specificità nella rappresentazione di entità extralinguistiche.

⁴ Per questo significato di "termini composti", vedi le considerazioni sul significato di "compound term" esposte nel par. 2.2 di questo articolo.

2. Termini e termini composti nelle definizioni degli standard ISO 2788:1986 e BS 8723-1/2:2005

2.1. Termini e concetti

Lo standard ISO 2788:1986, che afferma di occuparsi proprio delle relazioni *a priori* istituite mediante “associazioni mentali” tra i termini assegnati a un documento e altri termini implicati,⁵ definisce un termine di indicizzazione (“indexing term”) come la rappresentazione di un concetto per mezzo di un sostantivo o di un sintagma nominale; nell’ambito di un vocabolario controllato tale termine può essere denominato “termine preferito” o “termine non preferito”.⁶ Ciascun termine dovrebbe per quanto possibile rappresentare concetti semplici o unitari.⁷ Mentre in questo standard non è data una definizione di termine svincolata dall’attività di indicizzazione,⁸ il che come vedremo è fonte di ambiguità, nel più recente British Standard 8723:2005 “term” è descritto nella sua accezione più ampia, come una parola (“word”) o un sintagma (“phrase”) usata per etichettare (“label”) un concetto, ed è solo in seconda battuta riportato all’impiego nel contesto specifico dei thesauri, con la precisazione che nel contesto di un thesaurus i termini possono essere sia descrittivi (“descriptors”, o “pre-

ferred terms”), sia non descrittivi (“non-descriptors”, o “non preferred terms”).⁹

A differenza dello standard ISO, il British Standard definisce anche che cosa sono i “concetti”: si tratta di unità di pensiero (“unit of thought”) che esistono nella mente come entità astratte indipendenti dai termini che le rappresentano.¹⁰ Ciascun termine preferito o non preferito di un thesaurus deve rappresentare un solo concetto (“single concept”), e può essere formato da una sola parola (“single-word term”), o da più parole (“multi-word term”);¹¹ l’oggetto della significazione da parte dei termini formati da più parole sono tipicamente i “concetti complessi” (“complex concepts”), concetti che combinano due o più concetti semplici.¹² Viceversa, i termini che rappresentano i concetti complessi sotto forma di una unica o, nella maggior parte dei casi, di più parole, sono i termini preferiti scelti dopo che si è deciso che i concetti espressi dai termini stessi possono essere accettati nel thesaurus mantenendo l’espressione composta che presentano nella mente di chi ricerca e di chi costruisce il thesaurus, in base a considerazioni sulle finalità del thesaurus, sul tipo di documenti organizzati, ecc.¹³ Rispetto all’ISO 2788:1986, il BS 8723:2005 insiste frequentemente sul principio della significazione dei concetti da parte dei termini, fino a fondare quasi esclusivamente su di esso la natura stessa dei thesauri,¹⁴ cosicché i limiti dal punto di vista

⁵ Cfr. ISO 2788:1986 § 0 (“Introduction”): “Two kinds of inter-term relationships can be distinguished: a) syntactical or *a posteriori* relationships between the terms which together summarize the subject of a document. [...] b) those *a priori* or thesaural relationships between terms assigned to documents and other terms which, because they form part of common and share frames of reference, are present by implication. [...] Any of these mentally-associated terms might serve as a user’s approach to the subject index. [...] This International Standard is especially concerned with those *a priori* relationships which can be displayed in a thesaurus, where they then, in effect, add a second dimension to an indexing language [...]”.

⁶ ISO 2788:1986 § 3.4: “The representation of a concept, preferably in the form of a noun or noun phrase - NOTE: “[...] In a controlled indexing vocabulary, a term is designated either as a preferred term or as a non-preferred term”.

⁷ Cfr. ISO 2788:1986 § 7.1.1: “It can be stated, as a general rule, that terms should represent simple or unitary concepts as far as possible, and compound terms should be factored into simpler elements except when this is likely to affect the user’s understanding”. In ISO 2788:1986 § 6.1.1 sono indicate alcune categorie generali cui appartengono i concetti rappresentati dai termini di indicizzazione (“concrete entities”, “abstract entities”, “individual entities, or ‘classes-of-one’, expressed by as a proper nouns”).

⁸ Solo “indexing term” è definito; si vedano anche le definizioni di “indexing language”, “thesaurus”, “preferred term”, “non-preferred term”, “compound term”, “node label” in ISO 2788:1986 § 3.2-3.8.

⁹ BS 8723-1:2005 § 2.38: “term: word or phrase used to label a concept. NOTE: Terms in a thesaurus can be either descriptors or non-descriptors”.

¹⁰ BS 8723-1:2005 § 2.6: “concept: unit of thought. NOTE: Concepts can often be expressed in a variety of different ways. They exist in the mind as abstract entities independent of terms used to express them”. I nomi propri esprimono “entità”/“entità uniche” (cfr. BS 8723-2:2005 § 6.1.1.3: “unique entities expressed as proper names”); nello standard non ne è data definizione, ma dall’elenco di esempi proposti (luoghi, specifici oggetti, individui, organizzazioni, ecc.) possiamo presumere che si tratti di qualcosa di diverso dai “concetti”.

¹¹ BS 8723-2:2005 § 6.1.1.1: “Each preferred term included in a thesaurus should represent a single concept (or unit of thought). A concept may be expressed by a single-word term or a multi-word term”.

¹² BS 8723-1:2005 § 2.5: “complex concept: concept that combines two or more simpler concepts. Example: **human resource management** combines the idea of **people** with their usefulness as **resources** requiring **management**. NOTE: Complex concepts are sometimes expressed in a single word, but are more often conveyed by a multi-word term”; § 2.18: “multi-word term: term consisting in more than one word. Example: **human resource management**. NOTE: Multi-word terms typically label complex concepts and are admissible in a thesaurus as preferred terms”.

¹³ Cfr BS 8723-2:2005 § 7.3-§ 7.7.

¹⁴ Secondo il British Standard i thesauri sono vocabolari controllati in cui i concetti sono rappresentati da termini preferiti, organizzati formalmente in modo che le relazioni paradigmatiche (*a priori*) tra i concetti siano rese esplicite, il cui scopo è guidare l’indicizzatore e chi fa ricerche a selezionare lo stesso termine preferito o la stessa combinazione di termini preferiti per rappresentare un determinato argomento (BS 8723-1:2005 § 2.39: “thesaurus: controlled vocabulary in which concepts are represented by preferred terms, formally organized so that paradigmatic relationships between the concepts are made explicit [...] NOTE: The purpose of a thesaurus is to guide both the indexer and the searcher to select the same preferred term or combination of preferred terms to represent a given subject”; cfr. § 2.22: “paradigmatic relationship: relationship between concepts which is inherent in the concepts themselves. NOTE: Such rela-

teorico insiti nella visione anticipata nello standard del 1986 appaiono qui in modo ancor più evidente.

Non discuteremo in questa sede la risoluzione da parte degli standard del problema della significazione dei termini con un'adesione sostanzialmente acritica a una concezione psicologica e mentalista. Ci occuperemo invece dei problemi connessi con tale visione per ciò che concerne la decisione se accettare o meno all'interno del thesaurus un sintagma estratto da un documento.

2.2. Termini composti e concetti composti secondo l'ISO 2788:1986

Nello standard ISO 2788:1986, come per i termini di indicizzazione in generale, anche laddove sono trattati specificamente i termini composti il discorso è maggiormente incentrato sui termini, e, nel caso specifico dei termini composti ("compound terms"), sulla funzione delle loro parti componenti, che sul loro significato (i concetti composti/compleksi).¹⁵ In ogni caso né "compound term", né le espressioni come "complex subjects", "compound concepts", "complex notion", indicanti ciò a cui i "compound terms" si riferiscono, sono impiegati con un significato univocamente descritto. In particolare, le seconde sono fra loro intercambiabili, e non ricevono mai una definizione.¹⁶ Di "compound term" invece lo standard dà una definizione, che tuttavia non applica in ogni contesto.

Infatti, dapprima dice che un "compound term" è un termine di indicizzazione formato da più di una parola ("more than one word"),¹⁷ poi, che i termini composti possono

consistere in una sola parola ("single words", "single-word compounds"), nella quale sarebbero concatenate più parole frutto dell'individuazione di focus e differenza.¹⁸

In un altro punto si dice che "single words" e "compound terms" sarebbero ambedue oggetto di "semantic factoring", ma non è chiaro se in tale contesto "single words" stia per "single-word compound terms" o se le espressioni "single words" e "compound terms" siano in qualche modo contrapposte, poiché si riferiscono a entità differenti.¹⁹ In particolare, "compound term" è impiegato nel significato definito inizialmente di "termine di indicizzazione formato da più parole"? E come dovrebbe essere considerato, per esempio, "thermometers", "single word" o "compound term"?

In tutti questi casi l'ambiguità è generata dal fatto che lo standard non chiarisce preliminarmente che cosa intenda per "parola" (ingl. "word").

Considerato, inoltre, che la definizione di "termine composto" rimane iscritta nell'ambito dei sistemi di indicizzazione, non è chiaro in che senso nello standard si parli di accogliere o non accogliere termini composti nel thesaurus, dato che, quando siamo di fronte a un termine (di indicizzazione) composto la scelta di accoglierlo come preferito o non preferito nella sua espressione composta è già stata compiuta.²⁰ Ad accrescere le difficoltà di interpretazione è anche l'impiego in alcuni passi, accanto a "compound term", del semplice "term" e di "indexing term", in un modo per cui la relazione fra ciò che le tre espressioni significano non è perfettamente trasparente: "compound term" e "term" sembrerebbero infatti indicare parole o sintagmi *prima* del trattamento ai fini dell'indicizzazione.²¹

tionships are shown in a structured vocabulary, independently of any indexed document [...]. A paradigmatic relationship is sometimes known as an a priori relationship"). In BS 8723-2:2005 § 6.1.1.2 troviamo anche la lista, pur parziale, delle categorie in cui i concetti rappresentati dai termini preferiti di un thesaurus possono essere raggruppati in base a caratteristiche omogenee (cose e loro parti fisiche; materiali; attività o processi; ecc.). Inoltre in BS 8723-2:2005 § 5.1 è ricordato che l'uso frequente della parola "term" da parte dello standard non deve far dimenticare che lo scopo del trattamento dei termini è consentire l'individuazione dei concetti ad essi sottesi ("For reasons of practicality this standard refers frequently to 'terms'. But it should never be forgotten that the purpose of manipulating terms is to support retrieval of the underlying concepts").

¹⁵ Vedi ISO 2788:1986 § 3.7: "compound term: An indexing term [...] which can be factored morphologically into separate components, each of which could be expressed, or re-expressed, as a noun that is capable of serving independently as an indexing term. NOTE – The parts of the great majority of compound terms can be distinguished as follows: a) the focus, or head [...]. b) The difference or modifier [...]. The focus and the difference(s) may be written as separate words [...], or they may be concatenated into single words [...]."

¹⁶ Alcuni passi in cui si può trovare un riferimento ai concetti composti: ISO 2788:1986 § 7.1.1 ("complex subjects"), § 7.1.2 ("compound concepts"), § 7.1.3 a ("complex notion"), ecc.

¹⁷ ISO 2788:1986 § 3.4 ("note"): "An indexing term can consist of more than one word, and is then known as a compound term".

¹⁸ ISO 2788:1986 § 3.7: "The focus and the difference(s) may be written as separate words, as in 'dining rooms' and 'soup spoons', or they may be concatenated into single words, as in 'bedrooms' and 'teaspoons'; ISO 2788:1986 § 6.2.1: "Noun phrases belong to the category of compound terms, and occur in two forms: a) adjectival phrases, ex. MARINE BIRDS. This class also includes single-word compounds which can be factored morphologically into a noun plus a modifying difference having an adjectival function, ex. FOOTBALL, MOTORWAYS".

¹⁹ ISO 2788:1986 § 7.1.3: "This technique [Ndr., semantic factoring] can be applied to single words as well as to compound terms. A term which represents a complex notion is re-expressed in the form of simpler or definitional elements, each of which can also occur in other combinations to represent a range of different concepts".

²⁰ Vedi la definizione di termini di indicizzazione e di termini composti in ISO 2788:1986 § 3.7, § 3.4, e cfr. ISO 2788:1986 § 7.1.3 sulle tecniche di scomposizione, che comportano il non accoglimento di un "compound term" nel thesaurus.

²¹ Cfr. per esempio ISO 2788:1986 § 7.1.3 a ("A term which represents a complex notion is re-expressed in the form of simpler or definitional elements [...]."), § 7.1.4 ("In explaining the grounds for deciding whether a term should be retained in its pre-coordinated form or syntactically factored, it is necessary to distinguish the parts of a compound term through reference to their roles and interrelationships"), § 7.1.5 ("A small proportion of terms [Ndr., syncategorematic terms] encountered in indexing are not amenable to this kind of logical analysis [Ndr., morphological factoring]"), § 7.2.1 ("Compound terms should be retained as indexing terms in the following circumstances [...]."). L'impiego di "term" con significati diversi in contesti differenti contrasta anche con la dichiarazione che lo standard "is generally based on the concept of 'preferred terms'" (ISO 2788:1986 § 1.2 b).

Appare evidente, di fronte a questi esempi, che la sostanziale assenza di sistematicità dell'elaborazione teorica sottostante alle indicazioni operative per gli specialisti è causata da una serie di difficoltà nell'interpretazione delle indicazioni stesse. Possiamo dire che lo standard impiega espressioni come “compound term”, “morphological factor[ing]”, “head”/“modifier”, “derived” (detto di aggettivi derivati da sostantivi) ecc., che solo superficialmente richiamano il campo della linguistica, non essendo del resto possibile trarre dal riferimento a tale ambito di studi le dovute conseguenze per ciò che concerne la struttura dei thesauri. Fatte salve queste considerazioni, che approfondiremo nel par. 3.2, ci sembra plausibile che lo standard con “compound terms” voglia indicare sintagmi estratti dai documenti ai fini dell'indicizzazione, nei quali si possono distinguere diverse parti componenti ciascuna significativa in sé ai fini dell'accrescimento del thesaurus.

Del resto, già nel più recente British Standard si riscontra un parziale superamento dei problemi visti sinora riguardo alla definizione di “compound term”: i concetti complessi sono nella grande maggioranza dei casi espressi da “multi-word terms”, mentre i “single-word terms” che rappresentano concetti complessi sono citati solo marginalmente.²²

2.3. Concetti complessi nel BS 8723-1/2:2005

Abbiamo precedentemente citato (par. 2.1) la definizione data dal British Standard di “concept” come unità di pensiero. Nel definire i “concetti complessi” come concetti che combinano due o più concetti semplici, lo standard incorre in una aporia sul piano teorico: se il concetto è una “unità”, come possono esistere dei concetti passibili di ulteriore scomposizione? Se esistono concetti ulteriormente scomponibili, quali sarebbero le vere “unità” del pensiero? Inol-

tre, se si sostiene che anche una combinazione di più concetti è un concetto, allora “concetto” diventerebbe sinonimo di qualunque pensiero, compresi i testi, poiché la sintassi non avrebbe una funzione di composizione, essendo ogni combinazione di concetti concepita come una unità in sé – cosa cui peraltro né lo standard ISO, né il British Standard mostrano di credere, visto che riconoscono che tra i termini (o i concetti) possono essere istituite relazioni sintattiche capaci di produrre entità significative.²³

La definizione di “complex concept” è fonte di ambiguità anche se confrontata con quanto detto a proposito della significazione dei termini preferiti o non preferiti di un thesaurus, cioè che ognuno di essi dovrebbe rappresentare un solo concetto. Infatti, in che senso dovremmo intendere questa corrispondenza “uno-a-uno” tra termini e concetti se ammettiamo che ci sono termini, preferiti o non preferiti, formati da una sola, o, come più spesso accade, da più parole (“multi-word terms”), che rappresentano combinazioni di concetti, quali sono appunto i concetti complessi?²⁴ Nonostante la questione non sia affatto secondaria, dato che sarebbero i concetti complessi, in base alle esigenze di individuazione dei concetti da parte degli utenti, a essere accolti nel thesaurus sotto forma di un termine o di una combinazione di termini, oppure rifiutati,²⁵ il British Standard non viene in aiuto per chiarire quale sia la natura di questi particolari concetti. Vi si dice soltanto che possono presentare diversi gradi di complessità, e possono essere espressi da “multi-”/“single-word terms”,²⁶ nella sezione interamente dedicata ai concetti complessi (§ 7, *Complex concepts*) si parla solo dei termini composti da più parole che li esprimono: a partire da questi ultimi sono individuati focus e differenza, sempre poi indicando con concetti, o meglio, classi di concetti (“class of concepts”),²⁷ il riferimento delle parti componenti individuate.²⁸ Il continuo

²² Nel British Standard troviamo solo degli accenni a “single-word terms” che esprimono concetti complessi (in § 7.1 per esempio è spiegato come il concetto complesso espresso da un “single-word term” come **biodegradability** può essere trattato; cfr. anche BS 8723-2:2005 § 7.2: “A similar analysis can sometimes be applied to single-word terms representing complex concepts”); in § 7.1 è dichiarato che il trattamento dei concetti complessi è discusso a partire per lo più da “multi-word terms”, per ragioni pratiche, in quanto questi in genere rappresentano concetti complessi, e in rapporto ad essi è più semplice spiegare perché procedere alla scomposizione dei corrispondenti concetti complessi. Sul significato di “word” impiegato dal British Standard, vedi par. 3.2.

²³ ISO 2788:1986 § 0: “Two kinds of inter-term relationships can be distinguished: a) syntactical or *a posteriori* relationships between the terms which together summarize the subject of a document [...]; b) [...] *a priori* or thesaural relationships [...]”; l'ISO 2788:1986 parla anche di “scomposizione sintattica” di termini composti (§ 7.1.3 *b*, “syntactical factoring”; § 7.3, *Terms that should be syntactically factored*). BS 8723-2:2005 § 8.1.1: “Two kinds of relationships are commonly encountered between indexing terms. a) Relationships which exist only because the terms are used together in the context of a particular document, called syntagmatic relationships [...]. b) Relationships that are valid in almost all contexts, [...], known as paradigmatic relationships”; BS 8723-1:2005 § 2.36: “syntagmatic relationship: relationship between concepts that exist only because they occur together in a document being indexed. NOTE: [...] A syntagmatic relationship is sometimes known as an *a posteriori* relationship”.

²⁴ Su questo vedi anche il passo dell'ISO 2788 § 7.1.1 citato in nota 7, che prosegue: “Complex subjects should be expressed by combinations of separate terms, and these may be assigned as independent search keys in a post-coordinate system, or they may function as components of pre-coordinated index entries”.

²⁵ Cfr. BS 8723-2:2005 § 7.3, *Deciding whether or not to admit a complex concept*, e sgg.

²⁶ BS 8723-2:2005, § 7.1: “Concept range from the very simple to the very complex. At the simple end are notions such as silver or people; greater complexity is found in **Georgian silver teapots** or **human rights campaigners**. [...] Between the extremes are moderately complex concepts, which are very often conveyed by multi-word concepts, but can sometimes be expressed using just one term”.

²⁷ Qui e nel resto dell'articolo con “classe” ci limitiamo a ricalcare l'inglese “class”.

²⁸ Subito dopo aver esposto le considerazioni di carattere generale sui concetti complessi e sul loro trattamento (BS 8723-2:2005 § 7.1), lo standard spiega che cosa sono i “multi-word terms”, deputati tipicamente a rappresentare concetti complessi, e quali funzioni vi si possono individuare (*The nature of multi-word terms*, § 7.2).

passaggio da significante a significato senza che sia fatta davvero chiarezza sulla natura di ciò su cui si esercita l'attività di scomposizione rende difficoltosa l'interpretazione delle indicazioni fornite.

3. La scomponibilità dei sintagmi in base agli standard ISO 2788:1986 e BS 8723-1/2:2005

3.1. Scomposizione dei sintagmi e problema della significazione dei concetti complessi

Sia lo standard ISO, sia il British Standard, in realtà si concentrano più sui sintagmi accolti nel thesaurus nella loro espressione composta, che su quelli non accolti. Del resto, all'inizio dello standard ISO è dichiarato che la descrizione delle tecniche oggetto della trattazione generalmente si basa sulla nozione di "termine preferito".²⁹

Nonostante la limitata attenzione, l'operazione più difficile da giustificare dal punto di vista teorico è senz'altro la decisione di non accogliere nel thesaurus di un sintagma nella sua espressione composta, cioè la sua scomposizione in tanti termini quante sono le sue parti componenti.³⁰

Dal raffronto fra la descrizione della tecnica della scomposizione sintattica ("syntactical factoring"), che si applica ai "compound term", cioè ai sintagmi i cui focus e differenza possono essere a loro volta riespressi come termini indipendenti del thesaurus,³¹ e quella della scomposizione semantica ("semantic factoring"), della quale è sconsigliata la pratica a causa della perdita di precisione che comporta nel recupero delle informazioni,³² emerge un principio in base al quale la definizione di "termine composto" pare restringersi rispetto a quella data inizialmente dallo standard. "Compound terms" sarebbero sintagmi che possono essere

scomposti nei termini risultanti dall'individuazione di focus e differenza ("morphological factoring") *senza che ciò comporti una perdita di significato*, cioè in modo che sia sempre possibile comprendere il concetto composto oggetto della significazione. Il significato dei sintagmi ai quali può essere applicata la scomposizione sintattica è mantenuto sia nella loro risoluzione sintattica, sia nella loro espressione composta, che può essere accolta come non preferita in quanto punto di accesso utile (es. "Diritto tributario", USE "Tributi" + "Diritto", ovvero, in una stringa di soggetto relativa a un documento sul diritto tributario, "Tributi - Diritto").³³

Nello standard ISO, vedremo più avanti in dettaglio (par. 3.2), è frequente il riferimento alla *forma* dei "compound terms", che deve essere analizzata in ordine alla decisione se mantenere un sintagma nella sua espressione pre-coordinata, o se scomporlo nei termini che corrispondono alle sue parti componenti. Tuttavia, sia per questo standard, sia per il British Standard, che dedica molto spazio alle considerazioni di economia d'uso dei termini composti, si può dire che ciò che ispira l'individuazione delle circostanze in cui la scomposizione (sintattica) può avvenire è un criterio pragmatico: la naturalezza, o facilità, della rappresentazione sintattica di un sintagma, dovuta per esempio alla nostra abitudine alla forma invertita di alcune espressioni, cosicché esse sono per noi facilmente comprensibili. Questo è appunto il caso dei sintagmi le cui parti componenti sono fra loro nelle relazioni:

– parte (o proprietà)/tutto (e viceversa);³⁴

– azione transitiva/paziente dell'azione (e viceversa);³⁵

– azione intransitiva/soggetto agente (e viceversa).³⁶

Per esempio, un parlante la lingua italiana comprende con facilità che "Imprese - Gestione" significa "gestione di imprese", dal che è semplice trarre la regola generale per cui

²⁹ ISO 2788:1986 § 1.2: "This International Standard is [...] subject to the following restrictions: [...] b) it is generally based on the concept of 'preferred terms' [...]"

³⁰ Cfr. ISO 2788:1986 § 7.3.1: "It is more difficult to specify exact criteria for recognizing those kinds of compound terms which should be factored into separate components, each of which is then entered as a noun or a noun phrase in the thesaurus".

³¹ Cfr. ISO 2788:1986 § 7.1.3 *b*: "This technique is applied to compound terms, i.e. terms which are amenable to morphological analysis into separate components, each of which can be accepted as an indexing term in its own right".

³² ISO 2788:1986 § 7.1.3 *a*: "This technique can be applied to single words as well as to compound terms. A term which represents a complex notion is re-expressed in the form of simpler or definitional elements, each of which can also occur in other combinations to represent a range of different concepts. Example: thermometers could be expressed by a combination of three terms: TEMPERATURE + MEASUREMENT + INSTRUMENTS. This technique is not recommended. It can be employed only in post-coordinate indexing systems, and it should be applied only to terms in fringe subjects areas by agencies which need to limit the size of their indexing vocabularies. It is generally recognized that semantic factoring leads to a loss of precision in retrieval".

³³ Cfr. ISO 2788:1986 § 7.1.3 *b*: "If syntactical factoring is applied to a term in common use which may serve as a user's access point to the index, the term as a whole should be entered in the thesaurus as a non-preferred term, and a reference should be made to the elements used in combination".

³⁴ ISO 2788:1986 § 7.3.2.1 *a*: "A compound term should be factored if the focus refers to a property or part (including materials), and the difference represents the whole or possessor of that property or part: Examples: In the terms 1) AIRCRAFT ENGINES [...] the focal nouns, i.e. 1) ENGINES [...] represent parts or properties, and the differences 1) AIRCRAFT [...] represent their possessing wholes".

³⁵ ISO 2788:1986 § 7.3.2.2 *a*: "The name of a transitive action should not be modified by the name of the object on which the action was performed – Examples: In the terms: 1) OFFICE MANAGEMENT [...] the focal nouns 1) MANAGEMENT [...] represent transitive actions, and the differences 1) OFFICE [...] represent the patients of these actions. These terms should therefore be factored, and the parts re-expressed as separate nouns: 1) OFFICE + MANAGEMENT [...]"

³⁶ ISO 2788:1986 § 7.3.2.3 *a*: "The name of an intransitive action should not be modified by the name of the performer of the action – Examples: In the terms 1) BIRD MIGRATION [...] the focal nouns 1) MIGRATION [...] represent intransitive actions, and the differences 1) BIRD [...] represent the performers of these actions".

un sintagma il cui focus rappresenti un'azione transitiva e la differenza chi/ciò che subisce l'azione, deve essere scomposto; meno intuitiva sarebbe invece la rappresentazione sintattica "Automobili - Acquirenti" del sintagma "acquirenti di automobili", in cui la differenza "automobili" non indica una parte o proprietà del focus "acquirenti", ma è l'oggetto su cui si esercita l'azione di acquistare; purtuttavia, e solo apparentemente, è in rapporto paziente/azione transitiva con il focus, in quanto il focus non esprime un'azione, ma un soggetto che svolge un'azione ("acquirenti di automobili" = "coloro che acquistano automobili").

Come principio generale che guida la decisione di non accogliere un termine composto, la *possibilità d'uso* finisce dunque per imporsi sulla *forma*: del resto l'analisi formale, come vedremo in dettaglio nel par. 3.2, agisce a un livello inferiore della costituzione dei sintagmi, non rilevante ai fini dell'indicizzazione. Queste considerazioni di tipo pragmatico sono tutto ciò che riusciamo a trarre dallo standard ISO a proposito della scomponibilità dei sintagmi. Nel British Standard, in cui l'oggetto del trattamento sono i concetti complessi, a partire dai termini, in genere formati da più parole ("multi-word terms"), dei quattro risultati possibili del trattamento di un concetto complesso, solo due consistono nel non accogliere nel thesaurus il concetto complesso:

- il concetto complesso non è ammesso nel thesaurus, ma è rappresentato da un termine non preferito che punti a un unico termine meno significativo;³⁷
- il concetto complesso non è ammesso nel thesaurus, in quanto ricercato troppo poco frequentemente, non rilevante, o non adatto a essere rappresentato da un termine preferito o non preferito.³⁸

L'operazione di scomposizione ("splitting") di un concetto complesso, invece, non comporta la non ammissione del concetto nel thesaurus: il concetto è accolto nella sua inte-

grità, rappresentato attraverso una combinazione di termini che significano concetti più semplici, ed espresso attraverso un termine non preferito che punta a tale combinazione.³⁹ Possiamo dire che qui è reso in maniera esplicita quanto pareva di inferire dallo standard ISO a proposito della scomposizione (sintattica), cioè che il significato dei termini composti – il concetto complesso – è mantenuto nella loro risoluzione sintattica, nel caso in cui essi debbano essere scomposti. Fra i due standard si rileva tuttavia una differenza. Secondo l'ISO, la scomposizione ("semantic"/"syntactical factoring") porta a non accogliere il "compound term" nel thesaurus, e l'introduzione dello stesso come termine non preferito che punti alla combinazione dei termini prodotti dall'analisi delle parti componenti è solo un'eventualità giustificata dall'uso comune.⁴⁰

Seguendo il British Standard, invece, ogniquale volta si scompone un concetto complesso dovrebbe essere inserito nel thesaurus ("should be entered") un termine non preferito che rappresenti l'intero concetto, che rimandi alla corretta combinazione di termini indicanti concetti più semplici.⁴¹ Le condizioni favorevoli alla scomposizione individuate dal British Standard sono:

- l'eccessiva specificità del concetto complesso rispetto allo scopo primario del thesaurus;⁴²
- la previsione che il sintagma proposto per rappresentare il concetto complesso serva a indicizzare solo pochissimi documenti;⁴³
- più differenze specificano il focus del termine;⁴⁴
- il focus del termine rappresenta una proprietà, una parte o un componente della differenza.⁴⁵

Le procedure per la scomposizione si riducono allo "splitting" del concetto impiegando, nei casi più semplici, le parole che formano il corrispondente "multi-word term", oppure, in altri casi, parole che differiscono da quelle ricavabili a partire dal corrispondente "multi-word term".⁴⁶

³⁷ BS 8723-2:2005 § 7.3.1 c: "Reject the concept as a preferred term, but admit a non-preferred term pointing to a single broader preferred term. Examples: **camping holidays** USE **camping**; **thesaurus management software** USE **software**".

³⁸ BS 8723-2:2005 § 7.3.1 d: "Reject the complex concept, as too infrequently required, irrelevant or inappropriate to merit a preferred or non-preferred term".

³⁹ BS 8723-2:2005 § 7.3.1 b: "Admit the concept, representing it as a combination of two or more terms. In this case, a non-preferred term for the whole concept should be entered, pointing to the appropriate combination of simpler preferred terms. This option is often referred to as 'splitting' the concept. Examples: **camping holidays** USE **camping + holidays**; **thesaurus management software** USE **thesaurus management + software**".

⁴⁰ ISO 2788:1986 § 7.1.3 b.

⁴¹ BS 8723-2:2005 § 7.3.1 b; cfr. BS 8723-2:2005 § 7.4: "For complex concepts that do not justify a single preferred term but are likely to be sought by users, thesaurus entries of the following form should be provided (see 8.2.5 for more details)".

⁴² BS 8723-2:2005 § 7.3.3 a: "Splitting may be considered when the concept is quite specific and falls outside the core scope of the thesaurus. Inclusion of a large number of peripheral terms increases the bulk and complexity of the vocabulary without giving much retrieval benefit".

⁴³ BS 8723-2:2005 § 7.3.3 b: "If very few documents are likely to be indexed with the proposed term, inclusion as a preferred term may not be worthwhile, and splitting is an option to consider".

⁴⁴ BS 8723-2:2005 § 7.3.3 c: "The complex concept should be split if the term's focus is qualified by more than one difference. For example, **underwater cine cameras** should be split into **underwater cameras** and **cine cameras**".

⁴⁵ BS 8723-2:2005 § 7.3.3 d: "The following circumstances favour splitting: [...] d) The complex concept should usually be split if the term's focus represents a property, part or component of the difference – Examples: **aircraft engines**, **hospital floors**, **instrument reliability**, **soil acidity**".

⁴⁶ BS 8723-2:2005 § 7.4: "In simple cases, for example that of road safety, the concept may be adequately expressed using the words in the multi-word term, i.e. **roads + safety**. The justification for doing this in a particular thesaurus might be that any document dealing with road safety is likely to contain information relevant to both of the constituent terms and that somebody looking for information about roads is likely to judge an article on road safety as relevant. When a concept is split, the constituent concepts may be re-

Le condizioni che invece si oppongono alla scomposizione del concetto complesso sono:

- il termine che esprime il concetto complesso è tanto familiare nell'uso comune, o nell'ambito disciplinare coperto dal thesaurus, che rappresentarne gli elementi separatamente inficerebbe la comprensione;⁴⁷
- la scomposizione del concetto nelle sue parti componenti comporterebbe perdita di significato, o ambiguità;⁴⁸
- il concetto è rappresentato da un termine consolidato che è un nome proprio, o incorpora nomi propri;⁴⁹
- la differenza in un termine consolidato ha perso il suo significato originario;⁵⁰
- il concetto è rappresentato da un termine che contiene una differenza che suggerisce una somiglianza metaforica con un oggetto o un evento non irrelati;⁵¹
- le parti di un concetto una volta separate non esprimono l'intero concetto;⁵²
- il concetto è rappresentato da un termine in cui il focus ha un altro significato in assenza della sua differenza.⁵³

L'impressione generale che si ricava dalle indicazioni sulla scomposizione è che il British Standard lasci molta libertà, rispetto all'ISO 2788, di scomporre i sintagmi in base a valutazioni inerenti il loro utilizzo, nell'ottica sia di chi ricerca, sia di chi implementa il thesaurus.

Di conseguenza, la gamma di possibili esiti alternativi per uno stesso concetto complesso si allarga notevolmente. Ciò non sarebbe un problema, se alcune questioni fondamentali fossero state risolte dal punto di vista teorico: una di queste è la significazione dei concetti complessi da parte dei termini.

Prendiamo ad esempio il concetto "industria per la lavorazione del caffè".⁵⁴

Stando alle indicazioni dell'ISO, la differenza "per la lavorazione del caffè" non è con il focus in nessuna delle relazioni indicate a favore della scomposizione, operazione che comporterebbe lo smembramento del sintagma nel thesaurus, e quindi il suo mancato accoglimento.⁵⁵

Nel British Standard invece non troviamo alcuna prescrizione né a favore, né contro la scomposizione di questo concetto complesso, che comporterebbe l'espressione del suo significato attraverso la combinazione dei termini ricavabili a partire dalle sue parti componenti, e attraverso un termine non accettato formato da più parole ad essa collegato; non vi troviamo nemmeno la prescrizione di non accogliere affatto questo concetto complesso.⁵⁶

Decidiamo di scomporre "industria per la lavorazione del caffè", perché è troppo specifico rispetto allo scopo del thesaurus, oppure perché, verosimilmente, troppo pochi documenti saranno indicizzati impiegando il corrispondente termine formato da più parole, per cui riteniamo che non valga la pena includerlo nel thesaurus. Abbiamo anche verificato che non si determini alcuna delle circostanze individuate dallo standard che rendono sconsigliata la scomposizione.

La scomposizione può dare esito alle seguenti relazioni sostitutive: "Industrie per la lavorazione del caffè" USE "Lavorazione del caffè" + "Industrie", oppure "Industrie per la lavorazione del caffè" USE "Caffè" + "Lavorazione" + "Industrie", e ai libri che parleranno di industrie per la lavorazione del caffè saranno pertanto attribuite le stringhe di soggetto: "Lavorazione del caffè - Industrie", oppure: "Caffè - Lavorazione - Industrie".

La scelta tra le due alternative è legata alle priorità dell'agenzia: per esempio, la scelta di rappresentare il concetto

presented by preferred terms that differ from the words of the multi-word term. Often the constituent words need to be modified to convey the correct concept. For example, **rail safety** should not be split into **rails + safety**, but might be expressed as **railways + safety**. For **human resources** the combination of **people + resources** is probably more acceptable to users than **humans + resources**".

⁴⁷ BS 8723-2:2005 § 7.3.4 a: "a term for the concept has become so familiar in common use, or in the field covered by the thesaurus, that its expression as separate elements would hinder comprehension; Examples: **data processing, gross domestic product**".

⁴⁸ BS 8723-2:2005 § 7.3.4 b: "splitting the concept into its parts would lead to a loss of meaning, or ambiguity; Example: **plant food (food + plants)** could represent either food for plants, or plants as food".

⁴⁹ BS 8723-2:2005 § 7.3.4 c: "the concept is represented by an establishment term which is a proper name, or incorporates a proper name; Examples: **Boolean logic, United Nations**".

⁵⁰ BS 8723-2:2005 § 7.3.4 d: "the difference in an established term has lost its original meaning; examples: **lawn tennis, deck chairs, trade winds**".

⁵¹ BS 8723-2:2005 § 7.3.4 e: "the concept is represented by a term containing a difference suggesting a resemblance, as a metaphor, to an unrelated thing or event; Examples: **tree structure, wing nuts**".

⁵² BS 8723-2:2005 § 7.3.4 f: "the parts of the concept when separated do not convey the overall concept – thus it would be misleading to index the corresponding documents with terms representing the part concepts; Examples: **fire escapes** (a discussion of fire escapes might say nothing about fires or escaping), **swimming pools** (a discussion of swimming pools might have little information on swimming)".

⁵³ BS 8723-2:2005 § 7.3.4 g: "the concept is represented by a term in which the focus has another meaning in the absence of the difference, Examples: **artificial flowers, chocolate eggs** [...]".

⁵⁴ L'esempio è ricalcato su BS 8723-2:2005 § 7.3.1: "As the complexity increases, more choices have to be considered. Thus thesaurus management software could be represented by **thesauri + management + software** or by **thesaurus management + software**". Anche se, vista la forma in cui abbiamo espresso il concetto, non sarebbe necessario notarlo, con "industria" qui ci riferiamo a un'organizzazione, non a un settore produttivo.

⁵⁵ Cfr. ISO 2788:1986 § 7.1.4: "In explaining the grounds for deciding whether a term should be retained in its pre-coordinated form or syntactically factored, it is necessary to distinguish between the parts of a compound term through reference to their roles and inter-relationships".

⁵⁶ Cfr. BS 8723-2:2005 § 7.3.

complesso “industria per la lavorazione del caffè” per mezzo della combinazione di “Lavorazione del caffè” e di “Industrie” potrebbe essere motivata dalla necessità di distinguere la lavorazione del caffè da quella applicata ad altri materiali.⁵⁷ La scomposizione effettuata in questo modo causa tuttavia una perdita di significato: un libro che parla di industrie per la lavorazione del caffè potrebbe parlare poco o nulla dei procedimenti di lavorazione del caffè, o di industrie in generale; inoltre, la lavorazione del caffè non è un’attività significativamente diversa dall’attività di lavorazione applicata ad altri materiali, pur articolandosi secondo procedure ad essa peculiari. Analogo il discorso a proposito della rappresentazione di “industria per la lavorazione del caffè” per mezzo della combinazione dei tre termini “Caffè” + “Lavorazione” + “Industrie”, che potrebbe essere motivata per esempio dalla necessità di limitare la complessità del thesaurus,⁵⁸ oppure di venire incontro alle esigenze di un’utenza non specializzata.⁵⁹

In entrambi i casi, la rappresentazione sintattica non è di immediata comprensione, e il concetto complesso “industria per la lavorazione del caffè” non è ricostruibile con agio da parte di chi ricerca. Certo, il rischio di una perdita di significato è in qualche modo aggirato con la prescrizione di inserire un termine composto non preferito per l’intero concetto. Ciò potrebbe tuttavia risultare una forzatura qualora il termine non fosse davvero utile come ulteriore punto d’accesso, e si creerebbe un “rumore” eccessivo nel thesaurus: possiamo dire che il sintagma “Industrie per la lavorazione del caffè” è “di uso” tanto “comune” da poter essere inserito come ulteriore chiave d’accesso per “Lavorazione del caffè” + “Industrie”, o per “Caffè” + “Lavorazione” + “Industrie”? Non è piuttosto il termine che meglio esprime il concetto “industria per la lavorazione del caffè” (o, in altre parole, “impresa caratterizzata da un particolare tipo di organizzazione, dedicata alla lavorazione di un determinato materiale”), e pertanto dovrebbe essere accolto nel thesaurus come termine preferito “unico” che esprime il concetto complesso nella sua integrità?⁶⁰

Qui si apre una questione essenziale, che nessuno dei due standard analizzati affronta organicamente: il significato dei termini, o, nel caso particolare, la significazione dei concetti complessi da parte dei termini. Abbiamo già rilevato un’incongruenza a proposito della definizione di concetto complesso (cfr. par. 2.3); a questo punto viene spontaneo domandarsi in che senso possiamo dire che “indu-

stria per la lavorazione del caffè” è un “concetto complesso”, cioè un concetto che può essere scomposto in due o più concetti più semplici.

Lo stesso interrogativo vale per tutti i concetti complessi portati a esempio dallo standard: “Georgian silver teapots”, “human rights campaigners”, “human resources management”, “thesaurus management software” ecc. Tutti questi concetti sono detti “complessi”, ma in realtà si riferiscono a entità extralinguistiche uniche: “Georgian silver teapots” indica un particolare tipo di servizio da tè; “human rights campaigners” sono persone impegnate per il raggiungimento di un obiettivo particolare; “human resources management” indica l’attività gestionale applicata a un particolare oggetto; “thesaurus management software” indica programmi per elaboratori elettronici progettati appositamente per un’attività particolare; e così via. Nel caso di “industria per la lavorazione del caffè”, “per la lavorazione del caffè” nel suo insieme ha la funzione di esprimere una caratteristica (unitariamente, l’attività e il prodotto su cui si compie questa attività) del particolare tipo di organizzazioni indicato da “industria”.

I concetti che l’uso di tutti i sintagmi citati evocerebbe nelle menti degli utenti (e di chi costruisce thesauri) non si possono dire propriamente “complessi”: essi si riferiscono a entità uniche (es. servizi da tè) che presentano caratteristiche distintive (es. sono d’argento e sono stati realizzati all’epoca di un re Giorgio d’Inghilterra). Se concetti come questi nella lingua inglese, o nella lingua italiana, fossero espressi per mezzo di sostantivi in cui i parlanti non siano più in grado di distinguere le parti componenti, a ciascuna delle quali far corrispondere un concetto più semplice, nessuno sarebbe portato a procedere alla scomposizione (“syntactical factoring”, o “splitting”). Altrimenti si arriverebbe al paradosso che “industria per la lavorazione artistica del vetro” e “vetreria” sarebbero il primo un concetto complesso, il secondo no; eppure i parlanti la lingua italiana riconoscono che entrambi i sintagmi “industrie per la lavorazione artistica del vetro” e “vetrerie” designano il medesimo oggetto.

Non sono certo considerazioni sulla composizione del concetto indicato, bensì un criterio d’uso e di economia del thesaurus, a guidare l’accoglimento del termine “Vetrierie”, e il non accoglimento, se il caso introducendolo come non preferito, di “Industrie per la lavorazione artistica del vetro”. La rappresentazione del concetto di “vetreria” (o, è lo stesso, di “industria per la lavorazione artistica del vetro”) per mezzo di una combinazione di termini indicanti con-

⁵⁷ Cfr. BS 8723-2:2005 § 7.3.2: “If there is a large anticipated volume of relevant documents to be indexes, it may be important to discriminate between the complex concept and similar concepts. Where road safety occurs, for example, the same document collection may also deal with rail safety and/or transport safety more generally, as well as fire safety, safety in the home, etc. Many users want to discriminate between these different aspects of safety. Admitting all these terms gives an increased level of specificity, allowing greater precision in both indexing and searching”.

⁵⁸ Cfr. BS 8723-2:2005 § 7.3.3 a.

⁵⁹ Cfr. BS 8723-2:2005 § 7.3.2: “Another important factor is the expectations of the anticipated users. A term such **precoordinate indexes** could be acceptable and useful in a thesaurus to serve information retrieval specialists, but could be confusing if the thesaurus is intended for a more general audience”. Anche se poi si dice che l’opzione migliore in questo caso non è scomporre il concetto, ma accoglierne nel thesaurus il corrispondente termine formato da più parole, associandolo a uno dei termini corrispondenti a una delle sue parti (“In the latter case the best solution might be not to split the term into **precoordination + indexes**, but to choose option c), making it a lead-in term to the broader term indexes see 8.2.4”).

⁶⁰ BS 8723-2:2005 § 7.3.1 a: “Admit the concept, representing it with a single preferred term”.

cetti più semplici non può che apparire forzata, e il concetto indicato impossibile da ricostruire.

Il discorso è il medesimo per il concetto “industria per la lavorazione del caffè”: esso non è un concetto complesso, e non è scomponibile. Non essendo possibile in lingua italiana esprimerlo sotto forma di un “single-word term”, dovrà essere espresso per mezzo del sintagma formato da più parole:

Industrie per la lavorazione del caffè
BT Industrie alimentari
RT Caffè

Questa possibilità, va ricordato, non è peraltro esclusa né nel British Standard, né nello standard ISO, in cui non si trovano indicazioni esplicite contro l'accoglimento nel thesaurus di sintagmi come “industria per la lavorazione del caffè”. Anzi, lo standard ISO dice chiaramente che le procedure fissate non entrano nel merito della scelta tra forme diverse presenti nella stessa lingua a rappresentare gli stessi “concetti composti”: se nella lingua italiana “industria di telecomunicazioni” e “industria dei sistemi ed apparati per telecomunicazione”, che designano il concetto “industria che produce sistemi e apparati per la telecomunicazione”, sono egualmente attestati, non saranno considerazioni fondate su procedure standard bensì sull'uso di queste forme stesse a determinare la scelta dell'uno o dell'altro come termine accettato.⁶¹

In ogni caso, non avendo nessuno dei due standard stabilito punti saldi teorici ai quali ci si possa riferire per decidere se l'esito della scomposizione di un sintagma comporta o meno una perdita di significato, il riferimento ai concetti complessi come oggetto della significazione dei termini può indurre chi costruisce thesauri a considerare possibile la scomposizione anche quando la risoluzione sintattica del sintagma (o del concetto) sia incomprensibile. Gli stessi esempi riportati dal British Standard con riferimento ai concetti “biodegradability”, “human rights campaigners”, “camping holidays”, “thesaurus management software”, o “human resource management”, sono quantomeno discutibili, se si considera che lo stesso standard sconsiglia lo “splitting” del concetto complesso quando tale operazione porti a una perdita di significato (“loss of meaning”).⁶²

Conseguenze del riferimento ai concetti complessi analoghe a quelle che abbiamo visto a proposito dell'ammissi-

bilità nel thesaurus di sintagmi come “industria per la lavorazione del caffè”, si possono osservare anche in relazione al trattamento dei termini (o dei concetti) composti in cui più differenze modificano il focus.

Secondo entrambi gli standard, “industria chimica multinazionale” dovrebbe essere scomposto in componenti più semplici.⁶³ Tuttavia, “industria chimica multinazionale” si riferisce a industrie manifatturiere che producono prodotti chimici e che operano sul mercato mondiale attraverso aziende localizzate in diversi paesi; i libri che parlano di industrie chimiche multinazionali non parlano di industrie chimiche in generale e di industrie multinazionali in generale, ma di industrie chimiche *che sono anche multinazionali*, e, viceversa, di industrie multinazionali *che preparano prodotti chimici*: separando le due caratteristiche espresse per mezzo delle due differenze, l'oggetto della significazione si perderebbe irrimediabilmente, cioè non sarebbe più ricostruibile.

Le difficoltà nel reperimento dei documenti che trattano argomenti espressi mediante sintagmi caratterizzati da più differenze ci ha spinto a decidere di non applicare questa prescrizione degli standard nella nostra esperienza di costruzione del thesaurus ThESS.

3.2. L'analisi formale e l'equivoco del riferimento alla linguistica

Per capire quando secondo lo standard ISO e secondo il British Standard un sintagma è considerato scomponibile o meno, rimane da chiarire l'equivoco creato dal riferimento alla linguistica, in particolare alla morfologia.

Dal punto di vista della forma in cui si presentano, “autostrade”, “industrie calzaturiere”, “articoli per lo sport” secondo lo standard ISO sono tutti sintagmi nominali (“noun phrases”), appartenenti, in quanto tali, alla categoria dei termini composti. Il primo è un “single-word compound” in cui si possono individuare un focus rappresentato da un sostantivo e una differenza con valore aggettivale: appartiene per questo alla stessa classe dei sintagmi aggettivali (“adjectival phrases”) di cui fa parte anche “industrie calzaturiere”; il terzo è invece un sintagma nominale di tipo preposizionale (“prepositional phrase”).⁶⁴ In tutti e tre, siano le parole (“words”) che li compongono scritte separatamente o assemblate in una sola, è possibile individuare un focus

⁶¹ ISO 2788:1986 § 7.1.2: “The procedures reviewed in this International Standard are not concerned with the choice between such forms when more than one exists”.

⁶² BS 8723-2:2005 § 7.3.4 b.

⁶³ ISO 2788:1986 § 7.3.3: “[...These recommendations] also offer a basis for a standard treatment of compound terms where a given focus may be modified by more than one difference. If, for example, an indexing agency worked with policies which allow a phrase such as ‘aircraft engines’ in addition to the phrases ‘jet engines’ and ‘military aircraft’, the indexer would face difficulties when dealing with a complex subject such as ‘military aircraft jet engines’. In this kind of situation, the rules above offer a means of control by ensuring that the compound term is factored into separate and simpler components, i.e. ‘jet engines’ and ‘military aircraft’, each of which would be assigned to its own category in the thesaurus”. Cfr. BS 8723-2:2005 § 7.3.3 c: “The complex concept should be split if the term’s focus is qualified by more than one difference. For example, **underwater cine cameras** should be split into **underwater cameras** and **cine cameras**”. Cfr. anche BS 8723-2:2005 § 7.3.2.

⁶⁴ ISO 2788:1986 § 6.2.1: “An indexing term should preferably consist of a noun or a noun phrase. Noun phrases belong to the category of compound terms, and occur in two forms: a) adjectival phrases. Example: MARINE BIRDS. This class also includes single-word compounds which can be factored morphologically into a noun plus a modifying difference having an adjectival function, ex. FOOTBALL, MOTORWAYS. b) prepositional phrases. Example: HOSPITALS FOR CHILDREN”.

(identifica la classe di appartenenza del termine; in linguistica noto anche come “testa”, ingl. “head”) e una “differenza” (ingl. “difference”: specifica le caratteristiche in base alle quali il termine rappresenta un caso particolare della classe significata nel focus; in linguistica è detta anche “modificatore”, ingl. “modifier”).⁶⁵ L’individuazione delle diverse parti componenti può portare alla produzione di ulteriori termini del thesaurus, che possono essere inseriti come correlati a quelli considerati: per esempio, da “autostrade” ricaviamo “automobili”, da “industrie calzaturiere” ricaviamo “calzature”, da “articoli per lo sport” ricaviamo “sport”.⁶⁶ Cerchiamo ora di capire che cosa siano queste “parti” che compongono i sintagmi presi a esempio.

Nel par. 2.2, discutendo alcuni problemi circa il modo non perfettamente lineare in cui lo standard ISO impiega l’espressione “compound term”, abbiamo accennato al fatto che questo standard impiega “parola” (“word”) con riferimento all’aspetto dei termini composti, senza però mai darne una chiara definizione.

Anche il British Standard impiega “word” in espressioni come “single-word terms” o “multi-word terms” senza spiegarne il significato, ma ciò non costituisce un problema, poiché questo standard, spostando l’attenzione sul significato dei termini (i concetti), elimina quasi completamente il riferimento alla struttura grammaticale dei sintagmi che invece sembra ricorrere frequentemente nello standard ISO. Prendiamo in considerazione gli esempi di “compound terms” riportati nell’ISO 2788: “Dining rooms” e “Soup spoons”, formati da più parole scritte separatamente, e “Bedrooms” e “Teaspoons”, formati da più parole concatenate in una sola.

Ipotizzando che lo standard assuma una definizione molto generica di “word” come un’unità minima di significato di un enunciato,⁶⁷ quali e quante sono le “parole” che compongono questi termini? La scelta degli esempi citati, evidentemente gli uni contrapposti agli altri in base alla diversa forma (“dining room” vs. “bedroom”; “soup spoon” vs.

“teaspoon”), fa pensare che siano sempre due le “parole” implicate (*dining + room; bed + room; soup + spoon; tea + spoon*). Forse, di fronte a questi esempi, per un parlante di lingua inglese che cosa siano queste “parole” risulta intuitivamente chiaro. Ma a ben guardare la questione non è così semplice. Proviamo a spostarci nel contesto dei thesauri in lingua italiana. Quali e quante parole nel significato da noi ipotizzato formano “anticamera”, “cavatappi”, “guardaroba”?⁶⁸ E qual è la “testa” in “Cavatappi”, o in “Guardaroba”? Se parliamo di “morphological factoring”, riferendoci quindi alla morfologia, le unità in cui un termine composto può essere suddiviso non possono essere che i morfemi, le più piccole unità linguistiche dotate di significato di una parola o di un enunciato non ulteriormente scomponibili. Ma i morfemi corrispondono a quello che lo standard definisce come parte componente separata di un termine composto, che può essere espressa, o riespressa come un sostantivo in grado di costituire un termine indipendente di thesaurus? Viceversa, una definizione più intuitiva di “parola”, come quella che abbiamo ipotizzato sia stata assunta dallo standard, è coerente con la possibilità sancita dallo standard medesimo di riformulare le parti risultanti dal “morphological factoring” come ulteriori termini di un thesaurus se consideriamo morfemi quali ad esempio “per”, “da”, *anti-*, o *cava-*, che concorrono alla formazione dei termini sopra menzionati?

Si può rispondere a queste domande con la constatazione che solo apparentemente lo standard richiama nozioni proprie della linguistica, o della grammatica, come quelle di morfologia o di composizione delle parole. Proviamo a considerare per esempio quanto emerge dalla definizione di “semantic factoring” e “syntactical factoring”: pur esprimendo un concetto complesso, “thermometers” non può essere considerato un “compound term”, e quindi essere oggetto di trattamento, in quanto l’individuazione delle sue parti componenti non consente di ricavare sintagmi che possano funzionare come termini indipendenti del thesaurus senza che si verifichi una perdita complessiva di natura-

⁶⁵ Cfr. ISO 2788:1986 § 7.1.4: “The majority of compound terms, whether adjectival phrases or prepositional phrases, can be analysed into the following components: a) the focus (also known in linguistics as the ‘head’): the noun component which identifies the broader class of things or events to which the term as a whole refers. Examples: ‘CONCRETE’ in the adjectival phrase ‘REINFORCED CONCRETE’, ‘HOSPITALS’ in the prepositional phrase ‘HOSPITALS FOR CHILDREN’; b) the difference (also known as the modifier): that part of a compound term which refers to a characteristic, or logical difference, which when applied to the focus narrows its connotation and so specifies one of its subclasses. Examples: ‘REINFORCED’, which specifies a subclass of concrete in ‘REINFORCED CONCRETE’, ‘FOR CHILDREN’, which specifies a subclass of hospitals in ‘HOSPITALS FOR CHILDREN’” (cfr. anche § 7.1.5). BS 8723-2:2005 § 7.2: “The parts of most multi-word terms can be distinguished as follows: a) the focus or head, i.e. the noun component which identifies the general class of concepts to which the term as a whole refers. Examples: 1) the noun component ‘indexes’ in the term ‘printed indexes’; 2) the noun ‘hospitals’ in the prepositional phrase ‘hospitals for children’; b) the difference or modifier, i.e. one or more further components which serve to narrow the extension of the focus and so specify one of its subclasses. Examples: 1) the adjective ‘printed’ in the term ‘printed indexes’; 2) the preposition-plus-noun combination ‘for children’ in the term ‘hospitals for children’. A similar analysis can sometimes be applied to single-word terms representing complex concepts. Examples: term: **lawnmowers** – focus: **mowers**, difference: **lawn**; term: **bathrooms** – focus: **rooms**, difference: **bath**”.

⁶⁶ ISO 2788:1986 § 6.2.1: “Those parts of a compound term which function as differences [...] should be considered as potential sources of extra terms in a thesaurus. When a difference consists of an adjective, the noun from which the adjective was derived should be preferred as the extra candidate term. If these terms are accepted, the thesaurus should display reciprocal relationships between the extra term and the compound term as a whole”.

⁶⁷ “A combination of vocal sounds, or one such sound, used in a language to express an idea (eg. to denote a thing, attribute or relation), and constituting an ultimate minimal element of speech having a meaning as such; a vocable” (Voce: *Word*. In: *The Oxford English dictionary*. 2. ed., Oxford: Clarendon, 1989, vol. XX, p. 528).

⁶⁸ I termini proposti sono tutti esempi tratti dalla traduzione in lingua italiana dello standard ISO 2788:1986 § 3.7.

lezza della sua rappresentazione sintattica.⁶⁹ Del resto, dal punto di vista della costruzione di thesauri è importante solo la produzione di unità significative per chi ricerca, e questa finalità non sarebbe adempiuta se i thesauri fossero costruiti secondo una logica compositiva, a partire dalla struttura grammaticale dei sintagmi, senza riferimenti concettuali. Se questo fosse il caso, avremmo gerarchie come:

BT Birra-
 NT Birra
 NT Birrificio
 ecc.

o, ancora:

BT Auto-
 NT Automatismi
 NT Automazione
 NT Automobili
 (*ma non* NT Autostrade)

L'analisi in morfemi dovrebbe dunque appartenere a un livello differente, non toccato da chi costruisce thesauri. Essa agisce su un livello dei sintagmi inferiore rispetto a quello rappresentato dall'uso degli stessi da parte di chi scrive il documento e dai parlanti la lingua del documento; ed è quest'ultimo il livello di analisi che interessa a chi indicizza e a chi costruisce thesauri.

In ogni caso, il testo dello standard ISO, che non giustifica, riconducendolo alla dimensione dell'uso, l'impiego di nozioni quali "compound (terms)", "single-word compounds", "morphological factoring", "head"/"modifier", ecc., non aiuta a sciogliere l'ambiguità di fondo.

Può essere utile a questo punto riassumere il trattamento dei termini composti attraverso l'utilizzo non univoco di "factoring" da parte dell'ISO:

a) in associazione con "morphological" o "morphologically", "factoring" indica l'individuazione di focus e differenza dei termini composti: un'analisi di tipo formale che tuttavia, come abbiamo visto, non è "morfologica", ossia non scende al livello della struttura grammaticale dei sintagmi, e che, oltre a guidare la collocazione dei termini nelle gerarchie, può portare a produrre ulteriori termini per il thesaurus:

es.
 Autostrade
 BT Strade
 RT Automobili

 Industrie calzaturiere
 BT Industrie [in base al prodotto]
 RT Calzature

 Articoli per lo sport
 BT Strumenti [in base allo scopo]
 RT Sport;

b) in genere da solo, o accompagnato dall'aggettivo "syntactical", "factoring" indica l'analisi dei termini composti in diverse unità di significato ("analysing a term into separate meaning elements")⁷⁰, un'analisi di tipo contenutistico che porta a non accogliere nel thesaurus il sintagma analizzato, e a scomporlo nei termini risultanti dall'individuazione di focus e differenza, in alcuni casi mantenendolo anche nella sua espressione composta come termine non preferito collegato alla combinazione di termini risultato della scomposizione. Per distinguere le due diverse attività implicate nel significato di "factoring", proponiamo di chiamare "enucleazione" la produzione, da un sintagma accettato come termine, dei termini corrispondenti alle parti che lo costituiscono formalmente e non logicamente, e "scomposizione" l'operazione che comporta la non accettazione del sintagma come termine del thesaurus. Nel British Standard, invece, che si concentra esclusivamente sulla scomposizione ("splitting") dei concetti complessi, a partire da un'analisi formale dei termini ad essi corrispondenti, l'operazione di enucleazione di nuovi termini per il thesaurus a partire dalle parti componenti ciascun termine è implicitamente considerata come necessaria alla costruzione della rete di collegamenti dei termini.

In realtà l'esigenza di mantenere unite due diverse impostazioni, quella "ingegneristica" delle origini, in base alla quale anche i sintagmi formati da una sola parola in cui si individuino focus e differenza/e erano scomposti, e la tradizione dei "subject headings", che si rifà al linguaggio naturale, determina, da un lato, l'attenzione nei confronti della scomposizione di sintagmi formati da una sola parola (sintagmi come per esempio "football", "thermometer" nello standard ISO, "biodegradability" nel British Standard),⁷¹ sia la parzialità delle indicazioni sulla scomposizione, che non tengono conto del fatto che, anche quando focus e differenza di un termine composto siano fra loro nei tre rapporti descritti dallo standard, non è sempre possibile risolvere sintatticamente il termine analizzato.

Per esempio, il sintagma "motori automobilistici" estratto da un libro che parla dell'impiego dei motori automobilistici per produrre energia ad uso civile non può essere scomposto in "Automobili" + "Motori", anche se "automobilistici" apparentemente esprime un tutto rispetto a "motori". Il soggetto assegnato al libro in questo caso dovrà essere:

Motori automobilistici – Impiego per la produzione di energia
e non:
 Automobili – Motori – Impiego per la produzione di energia,

pena la incapacità del soggetto di rappresentare l'argomento del libro, che non parla di automobili, né di motori di automobili, ma dell'impiego per produrre energia di motori progettati appositamente per automobili (e non per esempio per aerei, o per navi).

⁶⁹ Cfr. ISO 2788:1986 § 7.1.3 *a e b*. Cfr. inoltre ISO 2788:1986 § 6.2.1 a proposito dei "single-word compounds" ("This class [i.e. adjectival phrases] also includes those single-word compounds which can be factored morphologically into a noun plus a modifying difference having an adjectival function").

⁷⁰ ISO 2788:1986 § 7.1.3.

⁷¹ V. par. 2.2, p. 5 e nota 22.

4. Conclusioni

Abbiamo rilevato nel testo dei due standard ISO 2788:1986 e BS 8723-1/2:2005 alcune ambiguità e contraddizioni per ciò che concerne la questione della scomponibilità di sintagmi incontrati nei documenti. Abbiamo mostrato le incongruenze nella definizione degli oggetti su cui si esercita l'attività di scomposizione di chi costruisce thesauri (termini composti per l'ISO 2788, concetti complessi per il British Standard), mettendo in dubbio che il principio della significazione dei concetti complessi da parte dei termini possa essere sostenuto senza incorrere in aporie dal punto di vista teorico, e nella pratica quando si tratta di decidere se accogliere o meno all'interno del thesaurus un sintagma incontrato nell'analizzare il contenuto di un documento.

Riguardo alla scomponibilità dei sintagmi, dalla nostra analisi emerge una maggiore prescrittività dello standard ISO rispetto al British Standard nel delineare i criteri – pragmatici, non teorici – in base ai quali l'operazione di scomposizione è giustificabile. Sebbene nello standard ISO ricorra spesso il riferimento alla forma dei sintagmi, la cui analisi è preliminare alla decisione di accoglimento nel thesaurus o di scomposizione, l'individuazione di solo pochi casi in cui la risoluzione sintattica dei sintagmi stessi è efficace determina due conseguenze: consente di raggiungere un certo grado di prevedibilità e una maggiore facilità di giustificazione delle scelte effettuate, e lascia a chi costruisce thesauri la libertà di accettare sintagmi in base a considerazioni sull'ammissibilità che possono partire da presupposti stabiliti dalle singole agenzie in base a particolari esigenze pratiche o punti di vista teorici.

Al contrario, il British Standard si sforza di colmare le lacune definitorie dello standard ISO, specie su che cosa sono termini e in che cosa consista il loro significato (i concetti), ed evita ambiguità come quelle generate, sempre nell'ISO 2788, da un esplicito riferimento alla linguistica; tuttavia, poiché in realtà non costruisce una teoria solida, si presta facilmente a essere utilizzato per giustificare scelte di scomposizione effettuate in base a considerazioni d'uso anche laddove la risoluzione sintattica dei sintagmi comporti l'impossibilità, per chi indicizza e per chi ricerca, di ricostruire il corrispondente concetto complesso. La perdita di significato in cui si dovesse in-

correre nel risolvere i concetti complessi nelle loro parti componenti è sì premessa da questo standard come circostanza in sé svantaggiosa per la scomposizione, ma i possibili esiti della scomposizione, lasciati esplicitamente all'arbitrio dei costruttori di thesauri, presentano problemi proprio dal punto di vista della comprensibilità dei concetti complessi stessi attraverso le diverse combinazioni che potrebbero rappresentarli.⁷² Scomporre un sintagma in due termini o in più termini insomma non può essere indifferente, o meglio, determinato solo da considerazioni sulle sue possibilità d'uso, poiché ne verrebbero meno la coerenza e la prevedibilità necessarie sia per chi utilizza il thesaurus, sia ai fini dell'interoperabilità fra i sistemi di agenzie differenti. Anche quando non ci si ponesse questi come obiettivi (il British Standard afferma che non è possibile né indispensabile raggiungere una coerenza assoluta nel trattamento dei concetti complessi, sebbene in alcune situazioni potrebbe essere utile una certa regolarità),⁷³ la scomposizione basata principalmente su considerazioni di economia d'uso inficerebbe la significatività dei termini del thesaurus e la loro possibilità di utilizzarli in ambiente pre-coordinato per esprimere gli argomenti dei documenti.

D'altra parte, nessuno dei due standard impedisce di accogliere in un thesaurus sintagmi che un'agenzia ritenesse significativi se mantenuti nella loro espressione composta, a condizione naturalmente che non si verifichi alcuna delle circostanze che ne renderebbero la scomposizione necessaria nell'economia del thesaurus. Ciononostante, i numerosi problemi che l'impianto teorico dei due standard presenta non possono essere ignorati, e richiedono, per essere superati, un serio ripensamento dei presupposti su cui si basa la costruzione di thesauri. È necessario comprendere quali siano le reali ragioni che debbano guidare chi deve costruire un thesaurus di fronte a sintagmi "non semplici" incontrati nei documenti. L'attribuzione di questa caratteristica a molti termini appare "naturale", ma quando è posta a confronto con la sua negazione si comprende quanto sia aleatoria. Probabilmente quelle che si dovrebbero indagare sono ragioni sistemiche, cioè interne al sistema di indicizzazione thesaurus-soggetti, che hanno a che fare, da un lato, con la necessità di costruire famiglie di termini coerenti, dall'altro con i vincoli posti dalla sintassi delle stringhe di soggetto.

⁷² Un esempio per tutti, tra quelli riportati dallo standard: **human resource management** per il quale esistono, come possibili opzioni di sintesi, **human resources + management**, o **people + resource management**, o ancora **people + resources + management** (BS 8723-2:2005 § 7.1).

⁷³ BS 8723-2:2005 § 7.6: "Absolute consistency in the admission of complex concepts is difficult to achieve and is not always necessary. [...] In some circumstances, however, consistent rules are helpful for thesaurus editors. For example, when a thesaurus is shared among several organizations [...]. Similarly, if thesaurus compilation is a networked project [...] some guidance is necessary". Cfr. invece ISO 2788:1986 § 7.1.1: "The procedures described below are offered as guidelines for indexers using controlled vocabularies who need criteria to ensure consistent practice in the treatment of compound terms. Within this context, it is considered that the establishment of such criteria constitutes an important step towards achieving consistency, and hence predictability, not only within a single indexing agency, but also among different agencies working together [...]".

Abstract

This article deals with the handling of compound terms and complex concepts to be included in a thesaurus, as it is accounted for by two of the most used standards for thesauri construction, ISO 2788 and BS 8723-1/2. The criteria for factoring or splitting are primarily discussed. In particular, the effectiveness or even the admissibility of syntactical restitution of the meaning of noun phrases in every case accepted by the standards is questioned. A number of ambiguities and contradictions are highlighted, due to the lack of a consistent theoretical justification. As both standards allow different approaches, a rethink of the conceptual basis of thesauri construction is recommended, with reference to the interrelationships between keywords and subject headings in an indexing system.