You Are What you Do. An Empirical Characterization of the Semantic Content of the Thematic Roles for a Group of Italian Verbs

Gianluca E. Lebani[1], Alessandro Bondielli and Alessandro Lenci[2]

University of Pisa
[1]gianluca.lebani@for.unipi.it
[2]alessandro.lenci@unipi.it

The nature of thematic roles is a central, yet controversial, issue both for models of linguistic competence as well as for models of sentence processing. McRae et al (1997b) proposed to treat thematic roles as verb-specific prototypes that can be empirically described by resorting to a modified version of the traditional feature norm paradigm. In this paper, we present the results of two norming experiments in which we extended this approach to incorporate the distinction between filler-inherent and verb-entailed features, the latter being further characterized on the basis of their association with one or more phases of the time course of the event.

In the first experiment, we asked to a group of speakers to list the prototypical characteristics of the fillers of two semantic roles, agent and patient, for a set of 20 Italian transitive verbs. We then manually annotated the collected features according to our classification of feature types. In the second experiment, we encouraged participants to list as many properties as possible to describe the verb roles with respect to three different time slots: before, during and after the event described by the verb takes place.

The collected data supports the claim in McRae et al (1997b) that thematic roles can be also treated as verb-specific concepts. The first experiment reveals differences between the agent and the patient roles, which instead disappears in the second experiment. The methodological novelty introduced in the latter is also able to highlight the interaction between the speakers' knowledge of verb

Journal of Cognitive Science 16-4: 401-430, 2015
Date submitted: 05/23/15 Date reviewed: 06/25/15
Date confirmed for publication: 08/16/15

roles and the temporal phases of the events.

Keywords: *thematic roles, verbs, semantic memory, feature norms*

1. Introduction

Notwithstanding its controversial definition, and the lack of consensus on its nature, the notion of thematic role (a.k.a. semantic role, theta role, case role) has been central to the study of verbal semantics in fields as distant as linguistics, lexicography, psycholinguistics and natural language processing.

Intuitively, a thematic role can be understood as the role played by an argument in the event described by a verb. In a psycholinguistic perspective, such a notion has been shown to play a role in sentence comprehension (for a review, see McRae & Matsuki, 2009) in a variety of ways: certain verbs and nouns appear to be able to prime the typical participants of the event they describe (Altmann & Kamide, 1999; Ferretti et al, 2001; Hare et al, 2009), given the instantiation of a given syntactic structure or the presence of certain grammatical cues (Traxler et al, 2000; Ferretti et al, 2001; 2007; Altman & Kamide, 2007); conversely, nouns appear to be able to prime verbs describing events in which they typically participate (McRae et al, 2005), to modulate the plausibility of the other participants to the event (McRae et al, 1998; Kami et al, 2003; Bicknell et al, 2010) and to help disambiguating the event described by a polysemous verb (Matsuki et al, 2011).

In computational linguistics, thematic roles have been widely exploited as a form of lexical semantic representation in computational semantic resources such as FrameNet (Baker et al, 1998) and VerbNet (Kipper-Schuler, 2005), and several models have been developed in order to automatically extract this information from corpora (Gildea & Jurafsky, 2002).

However, the theoretical status of this notion is far from clear. In the years a plethora of decompositional approaches replaced the original definition of thematic roles, notably introduced in the modern thinking by Fillmore (1968), according to which they should be seen as primitive verb-independent entities. Modern approaches, indeed, see thematic roles as what

Van Valin (1999) labeled “generalized semantic roles” (for a review, see Levin & Rappaport Hovav, 2005). The most influential and accepted model belonging to this family has been proposed by Dowty (1989; 1991), who represents thematic roles as groups of prototypical entailments imposed by a group of verbs on their arguments. The two basic roles in this model are that of “proto-agent” and “proto-patient”, each characterized in terms of features that do not constitute a set of jointly necessary and sufficient conditions, but rather a prototype. Each feature describes an entailment that contributes to the subjecthood or objecthood of an argument. For instance, the *“volitional involvement in the event or state”* and *“undergoing a change of state”* are described as properties that contribute to the agent and patient proto-roles, respectively (Dowty, 1991: 572). According to Dowty, given a verb, it is the argument bearing the most prototypical proto-agent features that is realized as the subject, while the argument bearing the most proto-patient features is realized as the object. A given argument of a verb can have more agent-like or patient-like properties than similar arguments of verbs not possessing any of the prototypical characteristics of a proto-agent or of the proto-patient. For instance, the subject of “to kill is a more prototypical agent than the subject of “to fear”, because the latter lacks the feature of “causing a change of state” that contributes to the characterization of the proto-agent role.

The main novelty of Dowty’s proposal is to assume that semantic roles have prototype structure, like other concepts do. By building on Dowty’s work, McRae and colleagues construct a theory in which a verb’s thematic roles are seen as concepts *“formed through the everyday experiences during which people learn about the entities and objects that tend to play certain roles in certain events”* (McRae et al, 1997: 141). In this view, a role is a concept composed by the set of typical properties possessed by its fillers. For example, the agent of the verb “to lie” can be partially described by the following characteristics: ‹insecure›, ‹feeling in danger› and ‹mean›. This is true even when it is impossible to label or identify the typical agent of a verb with a single lexical concept. In the case of “to lie”, for instance, we can identify a set of properties typically possessed by its typical agent even if we are not able to describe it with a label more specific than “human being”.

It should be stressed that the view on thematic roles exemplified by the McRae et al (1997b) model and the one advocated by Dowty's (1989; 1991) should be seen as complementary in their nature, rather than competitive. The latter is a way to model what Dowty calls "L-Thematic roles", that is general semantic roles like agent or patient that describe the general behavior of classes of verbs and govern the syntactic realization of their arguments, often relying on a set of theoretically-developed set of semantic primitives. The former model on the other side, aims at the description of the properties associated with verb-specific roles, at the same time showing how such a representation can be extracted from empirically derived semantic features, that is from the kind of knowledge that can be gained in a norming study. For instance, while the "generalized thematic roles" model allows us to know that the patient of a transitive verb is prototypically affected by a change of state, verb specific features associated to a given verb like "to move" allows us to further qualify such a change as a transition from a point in space to another.

2. Feature Norms

The feature norm paradigm has been widely employed in the psycholinguistic tradition to investigate the content of the human conceptual representation. In this paradigm, short linguistic descriptions ("features") are collected by asking a group of native speakers to describe a probe concept. Typically, a feature norm dataset has the form of a list of concept description pairs such as "chair" ‹has four legs› and "airplane" ‹flies›.

Following the route traced by Rosch & Mervis (1975), feature norms have been employed to design experiments (e.g. Aschcraft, 1978; Vigliocco et al, 2006), to build models of the human semantic memory (e.g. Collins & Loftus, 1975; Hinton & Shallice, 1991; McRae et al, 1997, Vigliocco et al, 2004; Storms et al, 2010) to account for patterns of category-specific disorders in anomic patients (e.g. Garrard et al., 2001; McRae & Cree, 2002; Vinson et al., 2003; Sartori & Lombardi, 2004) and to account for empirical phenomena like semantic priming (e.g. Cree, McRae, & McNorgan, 1999; Vigliocco, et al, 2004), conceptual combination (e.g. Hampton, 1997) and categorization (e.g. Smith et al, 1974).

Out of the psycholinguistic domain, the potentiality of this paradigm attracted some interest also from the natural language processing community, where scholars focused on the development of corpus-based method to build feature-like representations (e.g. Poesio et al, 2008; Baroni et al, 2010; Kelly et al, 2013) and on the integration of subject-elicited descriptions with other kinds of lexico-semantic knowledge (e.g. Barbu & Poesio, 2008; Andrews et al, 2009; Steyvers et al., 2011; Lebani & Pianta, 2012). Crucially, it is not necessary to treat such linguistics descriptions as faithful records of the human semantic memory. Rather, it is sufficient to assume, following Cree & McRae (2003), that they provide some kind of window into the mental conceptual representation, simply because it is exactly this representation that is used by the subject to cope with the experimental task. Moreover, as extensively discussed by McRae et al (2005), the linguistic nature of these description probably facilitates the collection of some kinds of information, like the typical parts of an object or its typical location, over kinds of information that are difficult to express linguistically, like the taste or smell of an object.

Prominent feature norms collections are nowadays available for a restricted set of languages, mainly English (Garrard et al, 2001; McRae et al, 2005; Vinson & Vigliocco, 2008; Devereux et al, 2013), Italian (Kremer & Baroni, 2011; Lebani, 2012; Montefinese et al, 2013; Lenci et al, 2013), Dutch (De Deyne et al, 2008) and German (Kremer & Baroni, 2011; Roller & Schulte im Walde, 2014). The methodology employed to build such collections, their lexical coverage and their structure are strongly influenced both by the theoretical framework of reference as well as by the goal of the study. A “reference” methodology, however, can be identified in the one employed by McRae et al (2005), mainly due to the influence that this work exercised on the relevant literature of the last decade.

In the canonical paradigm, description are collected by asking to group of speakers to write on paper few short descriptions for a set of target concepts. Such an approach proved to be extremely useful to investigate which concept properties are easier to recall, but the resulting distribution of concept properties appear very sparse and biased (on the topic, see Cree & McRae, 2003; Kremer & Baroni, 2011). As an example, roughly three quarter of the descriptions collected by McRae et al (2005) belong to their most frequent 7 description types (out of their total inventory of 27). Some scholars modified this paradigm by asking their participants to complete a form (e.g. Garrard et al, 2001; Devereux et al, 2013) or by asking them to answer to questions (e.g. Lebani, 2012). Moreover, some scholars move away from a strictly controlled "paper and pencil" setting and collected data by means of an online ad-hoc developed web interface (Frassinelli & Lenci, 2012; Lebani, 2012, Devereux et al, 2013) or resorting to crowdsourcing (Roller & Schulte im Walde, 2014).

Norms collections vary also in the number and types of described concepts, with a vast majority of collections focusing on the description of concrete object, with the notable exceptions of the collection by McRae et al (1997b) and that by Vinson & Vigliocco (2008), both reviewed in the next sections. To our knowledge, the most substantial datasets to-date available are the 866 concepts described in Devereux et al (2013), the 541 normed objects in McRae et al (2005), the 456 concepts described in Vinson & Vigliocco (2008), and the 425 concepts in the dataset by De Deyne et al (2008).

Another dimension of variation concerns the treatment of the raw description collected from the speakers. This process consists of two phases: the so called "normalization" phase, in which the relevant chunks of information contained in the raw descriptions are identified and extracted, and the classification phase, in which these chunks of information are labeled with the kind of information they convey. The normalization strategies employed in the literature can be organized into three classes: i.) virtually no normalization is performed (e.g. De Deyne et al, 2008); ii.) the linguistic descriptions are conformed to a phrase template (e.g. Garrard et al, 2001; McRae et al 2005; Kremer & Baroni, 2011; Lenci et al, 2013, Devereux et al, 2013); iii.) only the focal concept(s) of the description is retained (e.g. Vinson & Vigliocco, 2008; Lebani, 2012).

The factor that is most influenced by the theoretical background and goal of the work is the classification adopted to label the kind of information conveyed by each description. While few authors proposed an ad-hoc developed feature type classification (e.g. Garrard et al, 2001; Vinson & Vigliocco, 2008; Lebani & Pianta, 2010; Devereux et al, 2013), a de-facto standard is the classification proposed by Wu & Barsalou (2009), a modified version of which has been adopted by McRae et al (2005), Kremer & Baroni (2011), Frassinelli & Lenci (2012) and Lenci et al (2013).

2.1. The collection of verb-specific thematic role features

Even if the feature norm paradigm has been employed mainly to describe nominal concepts, some works showed how it can be successfully applied to gain knowledge on several aspects of the meaning of verbs. Out of the total of 456 concepts normed by Vinson & Vigliocco (2008), 287 denote actions, 217 of which expressed by verbs. In a previous work, the same scholars exploited these descriptions to develop and test a model of semantic representation, the FUSS model (Vigliocco et al, 2004), thus showing that objects and actions can be modeled in the same space. In these norms, there is no principled difference in the ways verbal and nominal features are collected: in both cases, subjects are simply asked to describe their meaning.

The approach exploited by McRae et al (1997b) is radically different: speakers were not required to characterize the event denoted by a verb, but to describe the characteristics of the prototypical agents and patients of the events described by it. Put differently, if the aim of the feature norm paradigm is to describe concepts, in the Vinson & Vigliocco's (2008) approach the target concepts are the events, while in the McRae et al's (1997b) one the target concepts are the thematic roles.

The dataset in McRae et al (1997b) is composed by the linguistic description produced by 32 subjects for the proto-agent and proto-patient role of 20 English transitive verbs. Each subject was asked to describe just one role for each target verb and was presented with a booklet in which every single role was explained with a simple sentence ("someone who is convicted" for the patient role of the verb "to convict") and was followed by ten empty lines. Crucially, the instructions explicitly stated that participants were not required to name the typical filler of a role (e.g. "judge" as the agent of the verb "to convict"), but the characteristics that are common to its typical fillers (e.g. ‹is old› or ‹is incorruptible›). Experiments were run in small groups of subjects in 45 minutes-long sessions.

Once collected, raw descriptions have been normalized to record synonymous (though not necessarily identical) descriptions under the same label, and to use different labels for different kinds of information. Overall, 3350 raw descriptions have been collected (1838 for the agent, 1692 for the patient), corresponding to 1573 different features (800 for the agent, 773 for the patient). On average, each subject produced 5.5 features per role, and each feature has been produced by 2.2. subjects. In the final dataset, all features with frequency below 3 have been discarded, thus reducing the number of different features to 445 (237 agents and 208 patients).

Even if the main reason why McRae and colleagues conducted this norming experiment was to create the stimuli for subsequent experiments enquiring the internal structure of thematic roles and their role in sentence comprehension, useful indications can be obtained even from a simple descriptive analysis of the resulting dataset. First of all, subjects clearly produced usable and consistent data, in line with the standard performance in a concrete object norming study, suggesting that the feature norm paradigm can be exploited for the description of thematic roles. In turn, this support the role-as-a-prototype-concept view assumed by these authors.

Moreover, the analysis of the relative frequency of consistent features (i.e. those produced by more than 3 subjects) over the total number of distinctive features failed to reveal a significant difference between agents and patients, at the same time showing a significant variability in the number of features per verb-role pair. What these data shows, according McRae et al (1997b), is that some thematic roles are better defined than others, a consequence of the well known fact that some verbs admit a more restrictive group of fillers for its agent or patient position than other verbs. Examples reported by these authors include highly consistent verb-role pairs like the agent role of the verb “to rescue” or the patient role of the verb “to teach”, as opposed to loosely defined pairs like the patient role of “to accuse” or “to serve”.

As pointed out by McRae and colleagues themselves, two kinds of descriptions can be associated to the prototypical fillers of a given verb thematic role: the characteristics that the fillers are expected to possess in order to be able to fit into a certain semantic role and those that a filler possesses as a result of its role in the event describe by the verb. Given the verb “to serve”, exemplar patient role features of the former kind can be ‹is rich› and ‹is powerful›, while exemplar of the latter type can be ‹is full›, which is acquired by patient as a consequence of the happening of the event described by the target verb. In the norming experiment reported in the next pages, we further elaborate this suggestion by separating the constant features from those that change as a consequence of the event, and by further characterizing the latter on the basis of the time course of the described event.

3. Characterizing Italian Verb-Specific Thematic Roles

The norming experiments described in these pages were inspired by the work by McRae et al (1997). Their goal is twofold: first of all, collecting human-generated thematic role features for a set of Italian verbs; secondly, trying to characterize the distinction between "entailed" and "permanent" features, that is the distinction between those characteristics that a participant possesses insofar as it takes part in an event, and those that belongs to the filler of the thematic role irrespectively of its participation to the event.

Examples of "permanent" features can be the fact that a person that is terrorizing someone else ‹is violent› or ‹is sadist›, which are property that this person possesses independently of its agentive role in a terrorizing event. "Entailed" features, like the fact that the patient of a terrorizing act ‹is cool› before the event takes place, but then ‹gets scared›, thus becoming ‹fearful›, are not inherent properties of a filler, but are rather properties that are true only if the filler participates in the event.

We modeled the time course of our events by distinguishing three simple phases (henceforth "time slot"), according to which the entailed features can be classified:

- B(EFORE)-features: this class includes those features possessed by the filler concept before the event described by the verb takes place; for instance, ‹is ill› for the patient of the verb "to cure".
- D(URING)-features: this class includes those characteristics possessed by the filler while the event described by the verb takes place; for instance, ‹speak› for the agent of the verb "to teach".
- A(FTER)-features: properties of these class are true of the filler after the event takes place; e.g. ‹feel fine› for the patient of the verb "to cure".

Crucially, entailed features are not bounded to belong to just one of our time slots. Some features can be true for two slots at the same time: the patient of the event “to terrorize” can be said to ‹be scared› both during the event as well as afterwards. Some entailed features might appear to be true for a thematic role of a verb regardless of the time course of the event. Examples are the feature ‹want revenge› referred to the agent role of the verb “to punish”, or the feature ‹is hungry› referred to the agent role of the verb “to eat”. It is worth remarking that even in this latter case the feature counts as entailed rather than permanent according to our definition. In fact, a feature is permanent only if it holds of an individual independently of its having a certain role in an event. A feature can hold of an individual throughout the event span, but it is entailed as long as the individual would not have it, had he not participated in the event.

3.1. A Collection of Italian Verb-specific Role Features

In order to collect the aforementioned semantic information, we tested two different empirical methodologies. We build a first collection of features by adapting the paradigm exploited by McRae et al (1997b) and subsequently annotating our features. One purpose of this experiment was to confirm McRae and colleagues’ main findings: subjects are able to provide consistent descriptions (i.e. lists of characteristics on which they agree) and that some verb-role pairs are easier to describe than others. In addition, the features were manually annotated according to the time slot they belong to, so as to be comparable with the data collected in the second experiment, in which the participants were explicitly asked to produce features for each time slot of each verb-role pair.

3.1.1. Method

SUBJECTS. Eleven native Italian-speakers participated voluntarily to the experiment (6 males, 5 females). Nine of them were students at the University of Pisa. On average, subjects were 22.82 years old (s.d. 2.48).

MATERIALS. The experimental stimuli consisted of 20 transitive verbs holding animate agents and patients. We first translated into Italian the stimuli used by McRae et al (1997b). In four cases, i.e. for the verbs “to evaluate”, “to investigate”, “to lecture” and “to instruct”, the most plausible Italian translation overlapped with one of the other target verbs. These cases were removed and replaced with other transitive verbs. The 16 target verbs borrowed from McRae et al (1997b) were: *accusare* (“to accuse”), *adorare* (“to worship”), *arrestare* (“to arrest”), *assumere* (“to hire”), *condannare* (“to convict”), *curare* (“to cure”), *divertire* (“to entertain”), *insegnare*[1] (“to teach”), *interrogare* (“to interrogate”), *intervistare* (“to interview”), *licenziare* (“to fire”), *punire* (“to punish”), *servire* (“to serve”), *soccorrere* (“to rescue”), *spaventare* (“to frighten”), *terrorizzare* (“to terrorise”). The four novel verbs were: *convincere* (“to convince”), *giudicare* (“to judge”), *incontrare* (“to meet”) and *uccidere* (“to kill”).

PROCEDURE. Features have been collected through an online experiment by exploiting an ad-hoc web interface. In order to access to the experiment, subjects received an invitation email. They were asked to complete two 30 minutes sessions, during which they had to describe 10 target concepts. Each session was preceded by extensive instructions. Prior to the first session, subjects were also asked to practice with two training verbs: *chiamare* (“to call”) and *colpire* (“to hit”). Overall, each subject produced features for all our 20 target verbs, presented in one of three random orders. Thus, each verb-role pair had been described by 11 subjects. Although relatively small, our sample size is comparable to the sample size (16 subjects per verb-role pair) tested by McRae et al (1997b).

Every participant was presented with a target verb per web page. On top of the page, there was a reminder with the task instructions, followed by a sentence of the form *“qualcuno CONDANNA qualcun altro”* (“someone convicts someone else”), containing the verb to be normed, in this example, the verb *condannare* (“to convict”). The page was then vertically split into two parts, one for the agent, the other for the patient. Each thematic role section was headed by a request of the form *descrivi chi condanna/chi è*

[1] Note that, differently from its English counterpart *to teach*, the verb *insegnare* always realizes its patient role as an indirect object: *insegna matematica ai bambini* (“she teaches math to children”). This verb has been kept to preserve the parallelism with McRae et al (1997b) as much as possible.

condannato (“describe who convicts/who is convicted”), followed by 10 empty lines. When each page was loaded, only one randomly chosen role was visible, and the subject has 90 seconds to fill as many visible lines as he could, after which the other role was shown and the previous one was hidden. The overall time available to describe a verb was 3 minutes, after which all roles disappeared and the subject were requested to press a button to move to the next target verb.

3.1.2. Normalization and Annotation

The raw linguistic descriptions produced by the participants have been inspected in order to remove all unwanted material like incomplete, ungrammatical or incomprehensible descriptions, as well as to filter out all cases in which a subject named the typical filler of a thematic role (e.g. ‹cop› as the typical subject of the verb “to arrest”).

The selected descriptions have then been normalized. The main goal of this crucial processing phase was to identify the meaningful chunks of information in each raw description and cluster together the synonymous features produced by different participants to count the frequencies of each feature. As an example, modals and auxiliaries have been stripped away, so that descriptions like ‹could be guilty› has been simplified and encoded as ‹is guilty›. Synonymous features produced by different participants were encoded by using the most recurrent linguistic form. Accordingly, then, if two participants produced the feature ‹is calm› and another subject produced the feature ‹is cool›, we treated all these descriptions as instances of the feature ‹is calm›, that is therefore recorded in our final dataset as being attested 3 times. Synonymous expressions produced by the same subject, however, were treated as redundancies and removed. Descriptions carrying multiple chunks of information, such as ‹is incompetent or unable›, were split into unitary features, in our example the features ‹is incompetent› and ‹is unable›.

The resulting features were then labeled by one of the authors as “permanent”, if they were describing an inherent property of the thematic role prototypical filler, or “entailed”, if they were a property that the filler acquired from fulfilling a role in the event. In the latter case, the annotator manually marked the associated time slot (or time slots, if the case). To test the inter-coder reliability of this annotation we randomly sample 100 annotated features and asked another native speaker to annotate them. The agreement between the annotators has been measured adopting Cohen's *k* (Cohen, 1960), an agreement coefficient that takes into account the agreement that can be achieved by chance. We obtained a *k* value of 0.709, a value that allows us to conclude that our annotation is fairly reliable, notwithstanding the lack of consensus on how agreement values should be interpreted (on the topic, see Artstein & Poesio, 2008).

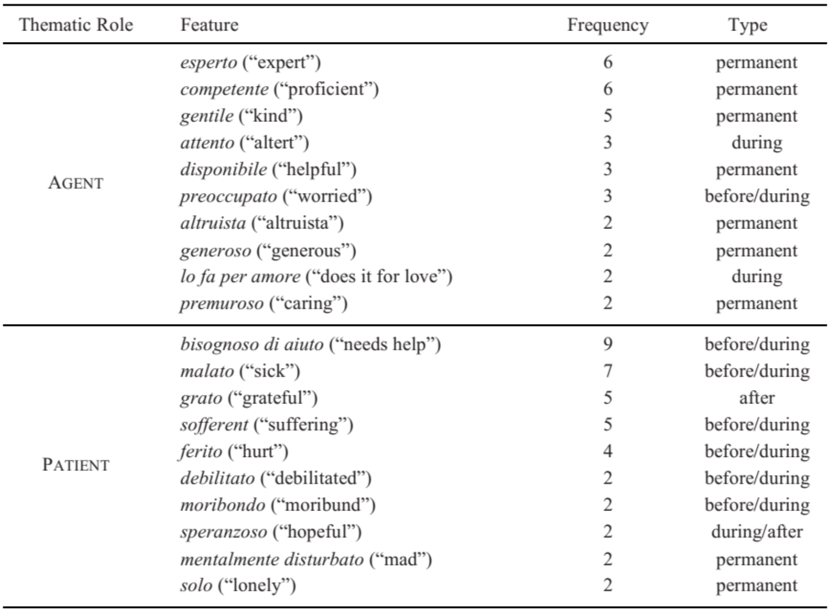
3.1.3. Results

The raw numbers of our dataset mirror those of the collections by McRae et al (1997b). Participants did not report any particular difficulty in listing the requested characteristics. Table 1 shows the consistent features collected for the verb *curare*, together with their production frequencies and their annotated feature types. On average, they produced 221.81 (s.d. 83.75) features each, 112.54 (s.d. 43.8) for the agent roles and 110.27 (s.d. 40.25) for the patient roles. For each role, they listed on average 5.66 (s.d. 2.42) features: 5.70 (s.d. 2.42) for the agent role and 5.61 (s.d. 2.42) for the patient one.

Overall, 2,249 raw descriptions were collected: 1,129 for the agent roles and 1,120 for the patient ones. From these, 1,514 distinct features were obtained 755 for the agent role and 759 for the patient. On the average, every feature has been produced by 1.48 (s.d. 1.02) participants: each agent feature by 1.49 (s.d. 0.98) subjects, each patient features by 1.47 (s.d. 1.06) subjects. 26.55% of the features were produced by more than 2 subjects (402): 212 for the agent role, 190 for the patient one.

Following the standard paradigm, we adopted a consistency threshold to filter out scarcely salient or idiosyncratic features. This practice is meant to maximize precision at the cost of recall, by balancing the effect of confounding variables such as autobiographical memory and limitations inherent in the experimental protocol. Due to the lower number of subjects that normed our stimuli, we opt for a threshold that is lower than the one employed by McRae and colleagues. Accordingly, we will consider

Table 1. Consistent features and production frequencies for the verb *curare* (“to cure”) in the first experiment



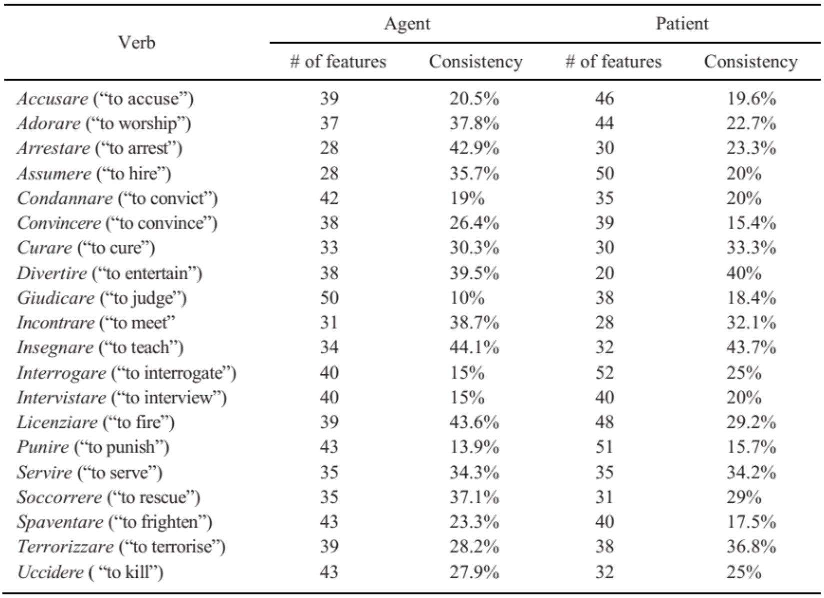
Thematic Role	Feature	Frequency	Type
AGENT	*esperto* (“expert”)	6	permanent
	competente (“proficient”)	6	permanent
	gentile (“kind”)	5	permanent
	attento (“altert”)	3	during
	disponibile (“helpful”)	3	permanent
	preoccupato (“worried”)	3	before/during
	altruista (“altruista”)	2	permanent
	generoso (“generous”)	2	permanent
	lo fa per amore (“does it for love”)	2	during
	premuroso (“caring”)	2	permanent
PATIENT	*bisognoso di aiuto* (“needs help”)	9	before/during
	malato (“sick”)	7	before/during
	grato (“grateful”)	5	after
	sofferent (“suffering”)	5	before/during
	ferito (“hurt”)	4	before/during
	debilitato (“debilitated”)	2	before/during
	moribondo (“moribund”)	2	before/during
	speranzoso (“hopeful”)	2	during/after
	mentalmente disturbato (“mad”)	2	permanent
	solo (“lonely”)	2	permanent

consistent all those features that are produced by at least 2 subjects (rather than 3), and calculate consistency as the percentage of consistent features over the number of distinct features.

Paired Student’s *t* tests failed to highlight any significant difference between agent and patient roles concerning both the percentage of distinct features over the total number of features ($t = -0.099$, $df = 19$, $p > 0.1$), and consistency ($t = 1.6$, $df = 19$, $p > 0.1$).

As in McRae et al (1997b), however, the distribution of consistent features in the verb-role pairs reported in Table 2 shows how some verb role pairs are clearly better defined than others. The most consistent agent roles are those held by the verbs *insegnare* (44.11%), *licenziare* (43.58%) and *arrestare* (42.85%); while the most consistent patient roles ones are those associated with the verbs *insegnare* (43.75%), *divertire* (40%) and *terrorizzare* (36.84%). On the contrary, the less consistent agent roles are those associated with the verbs *giudicare* (10%), *punire* (13.95%) and

Table 2. Number of distinguishing features and proportion of consistent features for each verb-role pair



Verb	Agent		Patient	
	# of features	Consistency	# of features	Consistency
Accusare ("to accuse")	39	20.5%	46	19.6%
Adorare ("to worship")	37	37.8%	44	22.7%
Arrestare ("to arrest")	28	42.9%	30	23.3%
Assumere ("to hire")	28	35.7%	50	20%
Condannare ("to convict")	42	19%	35	20%
Convincere ("to convince")	38	26.4%	39	15.4%
Curare ("to cure")	33	30.3%	30	33.3%
Divertire ("to entertain")	38	39.5%	20	40%
Giudicare ("to judge")	50	10%	38	18.4%
Incontrare ("to meet"	31	38.7%	28	32.1%
Insegnare ("to teach")	34	44.1%	32	43.7%
Interrogare ("to interrogate")	40	15%	52	25%
Intervistare ("to interview")	40	15%	40	20%
Licenziare ("to fire")	39	43.6%	48	29.2%
Punire ("to punish")	43	13.9%	51	15.7%
Servire ("to serve")	35	34.3%	35	34.2%
Soccorrere ("to rescue")	35	37.1%	31	29%
Spaventare ("to frighten")	43	23.3%	40	17.5%
Terrorizzare ("to terrorise")	39	28.2%	38	36.8%
Uccidere ("to kill")	43	27.9%	32	25%

intervistare (15%); the less consistent patient roles are held by the verbs *convincere* (15.38%), *punire* (15.68%) and *spaventare* (17.5%). Following McRae et al (1997b), this lack of consistency might reflect the fact that the latter thematic roles are associated with a broader range of fillers (i.e. almost anyone can be convinced, punished or frightened). However, the limited amount of data does not allow us to draw any firm conclusion on this point, that needs further investigation.

In order to analyze the distribution of the different feature types, we duplicated all features that were associated to more than one time slot (e.g. features that were labeled as jointly "before" and "during"), so that the features in the resulting dataset were associated only to one feature type (i.e. either "B", "D" or "A", in the case of the entailed features, "permanent" otherwise). This choice raised the number of consistent features to 435 (from 402 units). Two considerations lead us to this decision: first of all, we wanted our annotated data to be comparable with those collected in the second experiment (cf. below); more importantly, a

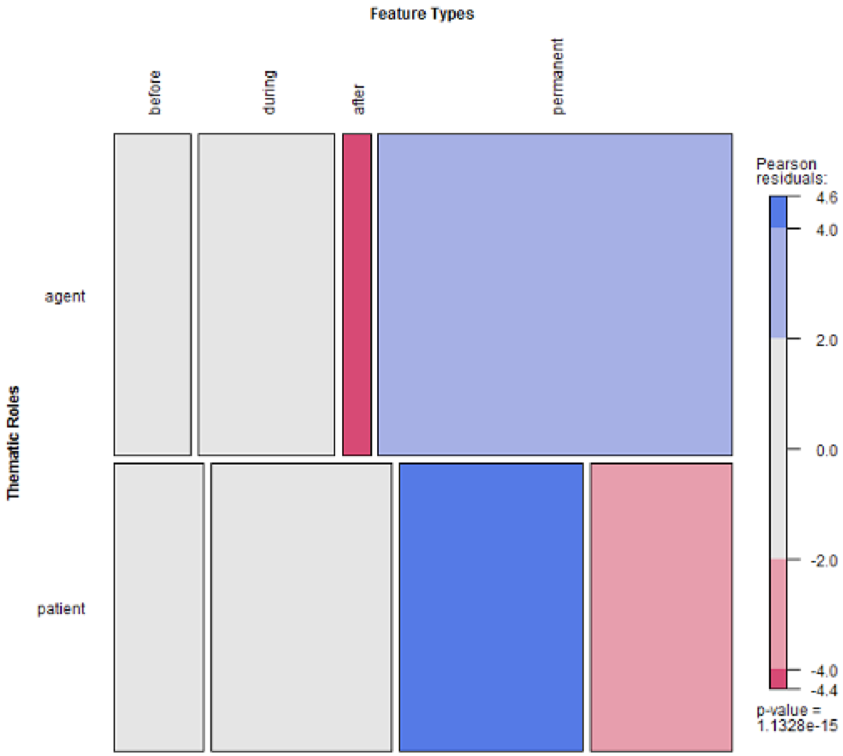


Figure 1. Cross-role distribution of the consistent (i.e. *frequency* ≥ 2) feature types in the first experiment

qualitative analysis of these cases revealed that the same features, when related to different time slots, were subject to a subtle meaning shift that we wanted to preserve. As an example, the feature ‹want revenge›, when referred to the agent role of the verb "to punish", can be true for each time slot. However, the following sentences suggest that the entailment relation between the verb and the feature somehow change in the different times slots: "he wanted revenge, that's why he punished him", as opposed to "even if he already punished him, he still wanted revenge".

A chi-square analysis revealed a significant difference in the distribution of feature types both in the whole dataset ($\chi^2 = 79.41$, df = 3, $p < 0.001$), as well as among the two groups of thematic roles ($\chi^2 = 72.69$, df = 3, $p < 0.001$). The mosaic plot (Meyer et al, 2006) in Figure 1 shows the details of this distribution. In this representation, the width of every rectangle is a function of the proportion of each feature type for each thematic role, while the height is proportional to the number of features produced for each thematic role. The analysis of the Pearson residuals is represented by the shadings of the cells: dark colored cells show a highly significant deviance from the expected values, while lighter shadings represent a medium-sized (still significant) deviance.

The Pearson residuals analysis highlighted a significant difference in the distribution of the permanent features in the two thematic roles: subjects tended to produce more permanent features for the agent role than for the patient role. Another significant different concerns the distribution of the entailed A-features, i.e. of those characteristics that the role fillers acquire after the event described by the verb took place: in this case, it is the patient role that shows a significantly higher association with this kind of features.

3.2. A Collection of Verb-specific Time-dependent Role Features

A well known issue affecting the feature norm paradigm concerns its natural bias towards some feature types (see Cree & McRae, 2003; Kremer & Baroni, 2010). Is this bias an effect of cognitive salience or is it due to the underrepresentation of some kinds of characteristics? Applied to our case: is the underrepresentation of the agent A-type features a consequence of their minor salience or do our subjects miss this kind of information? Moreover, in the first experiment subjects produce a high number of permanent properties (especially for agent roles), that are inherently associated with the fillers, independently of their participation in the event. Does this reflect the genuine content of the verb roles, or is it an effect of the way the data were collected, without any explicit mention of the event temporal phases? To investigate these issues, we collected a second group of norms by explicitly asking our participants to produce features for each time slot for each verb-role pair. In so doing, we pushed our participants to produce as many feature for type as possible. If some feature type are less represented than others, the bias recorded in the first experiment should manifest itself despite the change of the experimental paradigm.

3.2.1. Method

SUBJECTS. Eleven native Italian-speakers participated voluntarily to the experiment (7 males, 4 females). All of them were undergraduate students at the University of Pisa. On average, subjects were 24 years old (s.d. 1.42).

MATERIALS. The experimental stimuli consisted of the same 20 transitive verbs described in the first experiment.

PROCEDURE. The procedure exploited to collect features differed from the one used in the first experiment on a crucial aspect. The ten empty forms that composed each thematic role section (i.e. each of the two parts in which each page is vertically split) were substituted by three different subsections, one for each time slots. Every subsection was headed by a the name of the time slot, i.e. *Prima* ("before"), *Durante* ("during") and *Dopo* ("afterwards"). Each time slot was then followed by 5 empty forms. At every moment, all the time slots associated with a thematic role was visible to the speaker.

The time allotted was identical to that of the first experiment, i.e. 90 second for thematic role, in order to guarantee a proper comparison across the two settings. Such a choice is motivated by the view that the cognitive demand of the two settings is roughly comparable. In the first experiment, the relative simplicity of the task is counterbalanced by the vagueness of the experimental task. In the second experiment, the raise in task load due to higher number of questions addressed to the participant is counterbalanced by the more specific semantic information targeted by the slot-based setting.

3.2.2. Normalization and Annotation

The filtering and normalization phases were identical to those of the first experiment, with the difference that synonymous features produced by one subject were filtered out only if they were produced for the same verb-role-time slot. The resulting features were checked by one of the authors in order to identify and single-out possible "permanent" features produced by the participants.

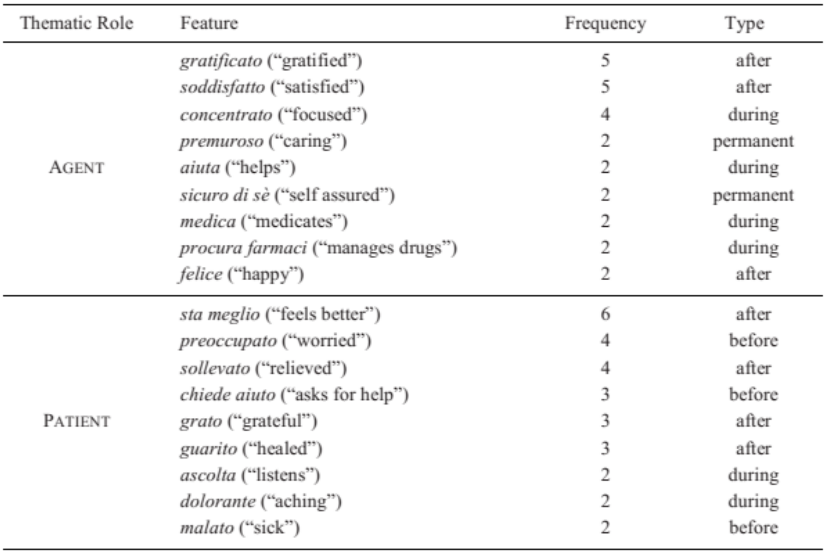
3.2.3. Results

Table 3 shows a subset of the consistent features collected for the verb curare, together with their production frequencies and their annotated feature types. Participants did not report any difficulty in listing features for each time slot. On average, each subject produced 284.63 (s.d. 119.38) features, 139.18 (s.d. 59.71) for the agent roles and 145.45 (s.d. 60.79) for the patient roles. Overall, 3,131 raw descriptions were collected: 1,531 for the agent roles and 1,600 for the patient ones. 1,039 features were produced for the B-type, 1,137 for the D-type and 955 for the A-type. From these, 2,278 distinct features were obtained: 1,127 for the agent role and 1,151 for the patient. 809 B-type distinct features were produced, 840 D-type, and 629 A-type. On the average, every feature was produced by 1.31 (s.d. 0.83) participants: each agent feature 1.29 (s.d. 0.79) times, each patient features 1.34 (s.d. 0.83) times. 19.09% of the feature were produced by more than 2 subjects (435): 196 for the agent role, 239 for the patient. 142 B-type consistent features were produced, 153 D-type and 140 A-type.

As in the previous experiment, a paired Student's *t* test failed to highlight any significant difference between agent and patient roles concerning the percentage of distinguishing features over the total number of features ($t = -0.965$, $df = 19$, $p > 0.1$), but the consistency of the patient role features is significantly higher than that of the agent features ($t = -2.33$, $df = 19$, $p < 0.05$). Again, the distribution of consistent features in the verb-role pairs reported in Table 4 shows that some verb-role pairs are clearly better defined than others, even though the range of variation is smaller than that of the first experiment.

A chi-square analysis revealed a significant difference in the distribution of feature types among the two groups of thematic roles ($\chi^2 = 42.8$, $df = 3$, $p < 0.001$). Crucially, we found also a significant difference in the distribution of the features types in the whole dataset ($\chi^2 = 65.77$, $df = 3$, $p < 0.001$), but the analysis of residuals showed that the only type of features whose frequency significantly deviates from the expected value are the permanents ones. In considering the distribution of the entailed feature types alone, we could not find any significant effect of the feature type ($\chi^2 = 0.72$, $df = 1$, $p > 0.1$).

Table 3. Consistent features and production frequencies for the verb *curare* ("to cure") in the second experiment



Thematic Role	Feature	Frequency	Type
AGENT	*gratificato* ("gratified")	5	after
	soddisfatto ("satisfied")	5	after
	concentrato ("focused")	4	during
	premuroso ("caring")	2	permanent
	aiuta ("helps")	2	during
	sicuro di sè ("self assured")	2	permanent
	medica ("medicates")	2	during
	procura farmaci ("manages drugs")	2	during
	felice ("happy")	2	after
PATIENT	*sta meglio* ("feels better")	6	after
	preoccupato ("worried")	4	before
	sollevato ("relieved")	4	after
	chiede aiuto ("asks for help")	3	before
	grato ("grateful")	3	after
	guarito ("healed")	3	after
	ascolta ("listens")	2	during
	dolorante ("aching")	2	during
	malato ("sick")	2	before

The major difference with respect to the first experiment is the strong reduction of permanent features, showing that the new method is better able to highlight the properties that a filler possess due to its involvement in the event. The mosaic plot in Figure 2 shows that subjects still produce a number of permanent properties. Like in the first experiment, permanent features are more for the agent role than for the patient role. This is a rather interesting effect, in that our data replicate the effect of the first experiment, even if we deliberately asked our subjects to focus on the different time slots of the events. On the other side, we could not replicate the significant difference in the number of A-type features produced for the agent and for the subject roles found in the first experiment.

Overall, the two experiments differ significantly both in the total number of distinct features ($t = -9.816$, $df = 19$, $p < 0.001$), and in the distribution of the different feature types ($\chi^2 = 139.73$, $df = 3$, $p < 0.001$). A paired t test, on the other side, failed to reveal a significant difference in the number of consistent features ($t = -1.295$, $df = 19$, $p > 0.1$). To investigate the

Table 4. Number of distinguishing features and proportion of consistent features for each verb-role pair

Verb	Agent		Patient	
	# of features	Consistency	# of features	Consistency
Accusare (“to accuse”)	58	17.2%	67	20.9%
Adorare (“to worship”)	43	11.6%	46	19.6%
Arrestare (“to arrest”)	57	22.8%	63	22.2%
Assumere (“to hire”)	53	18.9%	62	21%
Condannare (“to convict”)	57	15.8%	60	13.3%
Convincere (“to convince”)	53	17%	57	12.3%
Curare (“to cure”)	59	15.2%	59	28.8%
Divertire (“to entertain”)	56	17.9%	51	23.5%
Giudicare (“to judge”)	50	14%	59	15.2%
Incontrare (“to meet”	59	23.7%	55	21.8%
Insegnare (“to teach”)	59	13.6%	59	30.5%
Interrogare (“to interrogate”)	65	9.2%	68	19.1%
Intervistare (“to interview”)	61	13.1%	59	15.2%
Licenziare (“to fire”)	53	18.9%	56	21.4%
Punire (“to punish”)	54	24%	62	19.3%
Servire (“to serve”)	58	17.2%	52	23.1%
Soccorrere (“to rescue”)	63	20.6%	66	28.8%
Spaventare (“to frighten”)	42	19%	37	16.2%
Terrorizzare (“to terrorise”)	66	13.6%	61	21.3%
Uccidere (“to kill”)	61	24.6%	52	19.2%

the differences between the two experiments, in what follows we will analyze separately the features produced for the two roles.

The mosaic plot in Figure 3 compares the frequencies of the different types of feature obtained for the Agent role in the two experiments. While a paired t test failed to reveal a significant difference in the number of consistent features produced in the two experiments (t = 0.93, df = 19, p > 0.1), a chi-squared analysis shows a significant difference in the distribution of the different types ($\chi^2 = 94.6$, df = 3, $p < 0.001$). As shown by the mosaic plot, the difference in the number of A-type features obtained in the two experiment reaches statistical significance.

The mosaic plot in Figure 4 compares the frequencies of the different types of feature obtained for the Patient role in the first and in the second experiment. This difference is significant according to a chi-squared analysis ($\chi^2 = 64.4$, df = 3, $p < 0.001$). Differently to what happened for the Agent role features, in the second experiment our participants produced

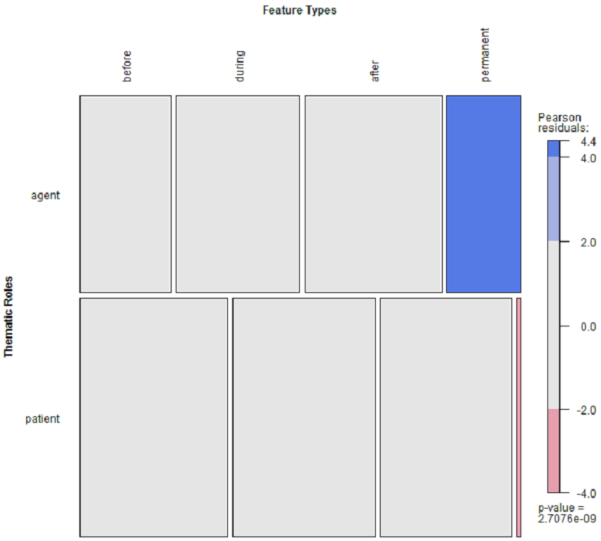


Figure 2. Cross-role distribution of the consistent (i.e. *frequency* ≥ 2) feature types in the second experiment

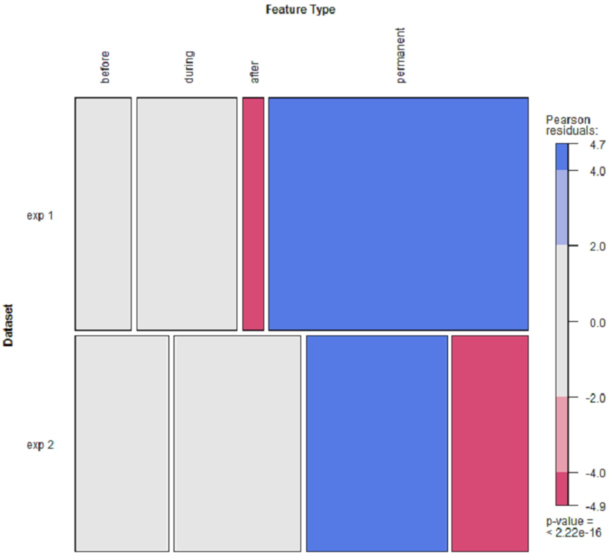


Figure 3. Distribution of the consistent (i.e. *frequency* ≥ 2) proto-Agent feature types in the two experiments

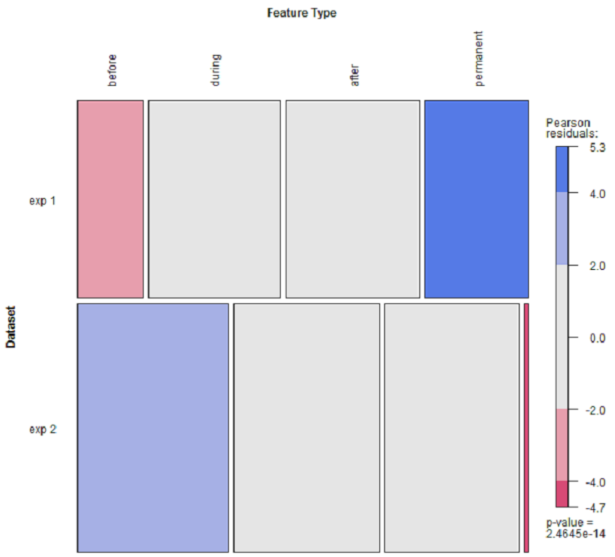


Figure 4. Distribution of the consistent (i.e. *frequency* ≥ 2) proto-Patient feature types in the two experiments

a significantly higher number of features for the patient role (t = -3.49, df = 19, p < 0.001). Besides the lower number of permanent features, which almost completely disappeared, in the second experiment our subjects produced a significant higher number of B-type features.

Taken together, we interpret these significant differences in the number of "entailed" features in the two collections as an effect of salience. When asked to freely describe a concept, our subjects tended to produce significantly more permanent features for the Agent role then for the Patient one, while the latter seems to be somehow more associated with what happens after the event described by the verb took place. When encouraged to list as many features as possible for each type, on the other side, our participants produced a significantly higher number of consistent descriptions for all those types that were underrepresented in the first experiment, so that the resulting distribution of "enhanced" property types looks rather balanced. This, in turn, suggests that all kinds of features are equally well-represented in our speaker's mind, and that the unbalanced distribution obtained in the first experiment is the consequence of the fact that some feature types for some thematic roles are more salient than others, and as a consequence more easily recalled, when subjects are not asked to focus on a specific event time slot.

4. Conclusion

In this paper, we presented the results of two norming experiments based on the verb-specific view of thematic roles proposed by McRae et al (1997b). Our major contribution in this paper is the presentation and evaluation of two different ways to collect and characterize the properties possessed by the prototypical fillers of a semantic role. We presented a feature type classification based on the distinction between filler-inherent and verb-entailed features, the latter further characterized on the basis of their association with one or more phases of the time course of the event. In the first experiment, features were collected independently of the event time slot. Since events unfold in time, we predicted that subjects might elicit different types of features for the various roles, when explicitly instructed to focus on the temporal phase before, during and after the event. The second experiment was therefore designed to test this hypothesis, which was indeed borne out by the feature analysis.

The significant difference in the number of A-type features associated with the two roles in the first experiment suggests the existence of a salience effect, according to which such features are more easily recalled to describe the patient role. This effect, in turn, can be linked to the idea that, at least for the kind of verbs we tested, the patient role is mainly characterized as the one affected by the verb. It would be tempting, moreover, to link this phenomenon to one of the prototypical properties listed by Dowty (1991) for the proto-patient, i.e. the fact the it “undergoes change of state”. However, as acknowledged by McRae et al (1997b) themselves, the relationships between the verb-specific thematic roles they propose and the higher order roles such as those widely exploited in the traditional literature on this topic is rather complex, and requires further study. In both experiments, moreover, we recorded a significant difference in the number of permanent features associated with each thematic role. Agent fillers seem to be more strongly associated with verb-independent characteristics, i.e. with properties that are required by the verb to select a filler as an appropriate agent, rather than with features that the fillers possess because of their participation in the event.

Such conclusions raise other questions that we plan to address in the future. For instance, it would be interesting to test to what extent the description of the "permanent" characteristics of the prototypical role fillers can benefit from the application of more fine grained feature type classifications, such as the one proposed by Wu & Barsalou (2009) or the one by Lebani & Pianta (2010). Possible extensions could take into account some of the several aspects of the verbal semantics neglected in this study, such as the aspectual properties of the described events. As an example, in these pages telic verbs, like "to arrest", and atelic verbs, like "to entertain", have been treated alike, while it is reasonable to suppose that such semantic characteristics can influence the behavior of a naïve speaker describing the temporal slots of a verb-role pair. Another test ground for our hypothesis could be the study of how the manipulation of the different feature types can influence sentence processing: is the thematic fit for the argument slot more sensitive to the "permanent" features of its candidate fillers the patient role? Are figurative interpretations triggered more easily by the violation of the "permanent" features or by the violation of the "entailed" ones? Do the different classes of "entailed" features play different roles in the interpretations of the main verb or of the entire sentence?

In conclusion, the main goal of this paper was to evaluate a framework for the description of the semantic content of verb-specific thematic proto-roles, in turn derived from the one proposed by McRae et al (1997b). At the same time, it highlighted some interesting systematic differences in the lexico-semantic representation of the agent and the patient roles, thus suggesting that the difference between these two roles can be partly characterized in terms of the salient features that are accessed by subjects in a norming experiment. Finally, the methodological novelty we have introduced in the second experiment proves to be suitable to probe into the interaction between the speakers' knowledge of verb roles and the temporal phases of the events.

Acknowledgements

This work was funded with the support of the PRIN grant 20105B3HE8 ”Word Combinations in Italian: theoretical and descriptive analysis, computational models, lexicographic layout and creation of a dictionary”, funded by the Italian Ministry of Education, University and Research. The authors would like to thank one anonymous reviewer for his/her helpful and constructive comments.

References

Altmann, G., & Kamide, Y. 1999. Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, 73, 247–264.

Altmann, G., & Kamide, Y. 2007. The real-time mediation of visual attention by language and world knowledge: linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, 57, 502–518.

Andrews, M., Vigliocco, G., & Vinson, D. P. 2009. Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, 116 (3), 463–98.

Artstein, R., & Poesio, M. 2008. Inter-Coder Agreement for Computational Linguistics. *Computational Linguistics*, 34 (4), 555–596.

Ashcraft, M. 1978. Property norms for typical and atypical items from 17 categories: A description and discussion. *Memory & Cognition*, 6, 227–232.

Baker, C. F., Fillmore, C. J. & Lowe, J. B. 1998. The Berkeley FrameNet Project. *Proceedings of COLING-ACL* ’98, 86–90.

Barbu, E., & Poesio, M. 2008. A Comparison of Feature Norms and WordNet. *Proceedings of the 4th Global Wordnet Conference* (GWC 2008), 56-73.

Baroni, M., Evert, S., & Lenci, A. 2008. Bridging the gap between semantic theory and computational simulations: *Proceedings of the ESSLLI 2008 Workshop on Distributional Semantics*, Hamburg: ESSLLI.

Baroni, M., Murphy, B., Barbu, E., & Poesio, M. 2010. Strudel: a corpus-based semantic model based on properties and types. *Cognitive Science*, 34 (2), 222–254.

Bicknell, K., Elman, J. L., Hare, M., McRae, K., & Kutas, M. 2010. Effects of event knowledge in processing verbal arguments. *Journal of Memory and Language*, 63 (4), 489–505.

Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46.

Collins, A. & Loftus, E. 1975. A spreading-activation theory of semantic processing. *Psychological review*, 82, 407–428.

Cree, G. & McRae, K. 2003. Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General*, 132, 163–201.

Cree, G. S., McRae, K., & McNorgan, C. 1999. An attractor model of lexical conceptual processing: Simulating semantic priming. *Cognitive Science: A Multidisciplinary Journal*, 23 (3), 371–414.

De Deyne, S., Verheyen, S., Ameel, E., Vanpaemel, W., Dry, M. J., Voorspoels, W., & Storms, G. 2008. Exemplar by feature applicability matrices and other Dutch normative data for semantic concepts. *Behavior Research Methods*, 40 (4), 1030–1048.

Devereux, B. J., Tyler, L. K., Geertzen, J., & Randall, B. 2014. The Centre for Speech, Language and the Brain (CSLB) concept property orms. *Behavior Research Methods,* 46(5), 1119-1127.

Dowty, D. 1989: On the Semantic Content of the Notion "Thematic Role". In G. Chierchia, B. H. Partee, & R. Turner (Eds.) *Properties, Types, and Meanings*, vol. II, 69–130. Dordrecht: Kluwer.

Dowty, D. 1991. Thematic Proto-Roles and Argument Selection. *Language*, 67 (3), 547–619.

Ferretti, T. R., Kutas, M. & McRae, K. 2007. Verb aspect and the activation of event knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 182–196.

Ferretti, T. R., McRae, K., & Hatherell, A. 2001. Integrating Verbs, Situation Schemas, and Thematic Role Concepts. *Journal of Memory and Language*, 44 (4), 516–547

Fillmore, C. J. 1968. The Case for Case. In E. Bach & R. T. Harms (Eds.) *Universals in Linguistic Theory*, 1–88. New York :Holt, Rinehart, and Winston.

Frassinelli, D. & Lenci, A. 2012. Concepts in context: Evidence from a feature norming study. *Proceedings of the Annual Meeting of the Cognitive Science Society* (CogSci 2012), 1566-1571.

Garrard, P., Lambon Ralph, M. A., Hodges, J. R., & Patterson, K. 2001. Prototypicality, distinctiveness, and intercorrelation: Analyses of the semantic attributes of living and nonliving concepts. *Cognitive Neuropsychology,* 18 (2), 125–174.

Garrard, P., Ralph, M. A. L., Hodges, J. R., & Patterson, K. 2001. Prototypicality, distinctiveness, and intercorrelation: Analyses of the semantic attributes of living and nonliving concepts. *Cognitive Neuropsychology,* 18 (2), 125–174.

Gildea D., & Jurafsky, D. 2002. Automatic labeling of semantic roles. *Computational Linguistics*, 28 (3), 245-288

Hampton, J. A. 1979. Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 18 (4), 441– 461.

Hare, M., Jones, M., Thomson, C., Kelly, S., & McRae, K. 2009. Activating event knowledge. *Cognition*, 111(2), 151–167.

Hinton, G. & Shallice, T. 1991. Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological review*, 98, 74–95.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. 2003. The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49, 133–156.

Kelly, C., Devereux, B. J., & Korhonen, A. 2013. Automatic Extraction of Property Norm-Like Data From Large Text Corpora. *Cognitive Science,* 38 (4), 638-682.

Kipper-Schuler, K. 2005. V*erbnet: A Broad-Coverage, Comprehensive Verb Lexicon*. Ph.D. thesis, University of Pennsylvania.

Kremer, G., & Baroni, M. 2010. Predicting Cognitively Salient Modifiers of the Constitutive Parts of Concepts. *Proceedings of the 2010 Workshop on Cognitive Modeling and Computational Linguistics*, 54–62.

Kremer, G., & Baroni, M. 2011. A set of semantic norms for German and Italian. *Behavior Research Methods*, 43 (1), 97–109.

Lebani, G. E. 2012. *STaRS.sys: designing and building a commonsense-knowledge enriched wordnet for therapeutic purposes.* Ph.D. Thesis, University of Trento.

Lebani, G. E., & Pianta, E. 2010. A Feature Type Classification for Therapeutic Purposes: a preliminary evaluation with non-expert speakers. *Proceedings of the 4th Linguistic Annotation Workshop (LAW IV)*, 157–161.

Lenci, A., Baroni, M., Cazzolli, G., & Marotta, G. 2013. BLIND: A set of semantic feature norms from the congenitally blind. *Behavior Research Methods*, 45 (4), 1218–1233.

Levin, B., & Rappaport Hovav, M. 2005. *Argument Realization*. Cambridge: Cambridge University Press.

Matsuki K., Chow T., Hare M., Elman J. L., Scheepers C., McRae K. 2011. Event-based Plausibility Immediately Influences On-line Language Comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 37 (4), 913–934.

McRae, K., & Cree, G. S. 2002. Factors underlying category-specific semantic deficits. In E. M. E. Forde & G. Humphreys (Eds.), *Category-specificity in mind and brain*. East Sussex, UK: Psychology Press.

McRae, K., & Matsuki, K. 2009. People Use their Knowledge of Common Events to Understand Language, and Do So as Quickly as Possible. *Language and*

Linguistics Compass, 3 (6), 1417–1429.

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. 2005. Semantic feature production norms for a large set of living and nonliving things. B*ehavior Research Methods*, 37 (4) , 547–559.

McRae, K., De Sa, V., & Seidenberg, M. 1997a. On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, 126, 99–130.

McRae, K., Ferretti, T. R., & Amyote, L. 1997b. Thematic Roles as Verb-specific Concepts. *Language and Cognitive Processes*, 12 (2/3), 137–176.

McRae, K., Hare, M., Elman, J. L., & Ferretti, T. R. 2005. A basis for generating expectancies for verbs from nouns. *Memory & Cognition*, 33 (7), 1174–1184.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. 1998. Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38, 283–312.

Meyer, D., Zeileis, A., & Hornik, K. 2006. The strucplot framework: Visualizing multiway contingency tables with vcd. *Journal of Statistical Software,* 17(3).

Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. 2013. Semantic memory: A feature-based analysis and new norms for Italian. *Behavior Research Methods*, 45(2), 440–461.

Poesio, M., Barbu, E., Giuliano, C., & Romano, L. 2008. Supervised relation extraction for ontology learning from text based on a cognitively plausible model of relations. *Proceedings of the 3rd Workshop on Ontology Learning and Population*, 1-5.

Roller, S., & Schulte im Walde, S. 2014. Feature Norms of German Noun Compounds. *Proceedings of the 10th Workshop on Multiword Expressions (MWE 2014),* 104–108.

Rosch, E. & Mervis, C. 1975. Family resemblances: Studies in the internal structure of categories. *Cognitive psychology*, 7, 753–605.

Sartori, G., & Lombardi, L. 2004. Semantic relevance and semantic disorders. J*ournal of Cognitive Neuroscience,* 16 (3), 439–452.

Smith, E. E., Shoben, E. J., & Rips, L. J. 1974. Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review,* 81 (3), 214–241.

Steyvers, M., Smyth, P., & Chemuduganta, C. 2011. Combining Background Knowledge and Learned Topics. *Topics in Cognitive Science,* 3 (1), 18–47.

Storms, G., Navarro, D., & Lee, M. 2010. *Acta psychologica special issue on formal modeling of semantic concepts*, 133 (3), 213–304.

Traxler, M. J., Fodd, D. J., Seely, R. E., Kaup, B. & Morris, R. K. 2000. Priming in sentence processing: intralexical spreading activation, schemas, and situation models. *Journal of Psycholinguistic Research*, 29, 581–595.

Van Valin, R. D., Jr. 1999. Generalized Semantic Roles and the Syntax-Semantics Interface. In F. Corblin, C. Dobrovie-Sorin, & J.-M. Marandin (Eds.) *Empirical Issues in Formal Syntax and Semantics, 2,* 373–389. The Hague: Thesus.

Vigliocco, G., Vinson, D., Lewis, W., & Garrett, M. 2004. Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology*, 48, 422–488.

Vigliocco, G., Warren, J., Siri, S., Arciuli, J., Scott, S., & Wise, R. 2006. The role of semantics and grammatical class in the neural representation of words. *Cerebral Cortex,* 16, 1790–1976.

Vinson, D. P., & Vigliocco, G. 2008. Semantic feature production norms for a large set of objects and events. *Behavior Research Methods*, 40 (1), 183–190.

Vinson, D., Vigliocco, G., Cappa, S., & Siri, S. 2003. The breakdown of semantic knowledge: Insights from a statistical model of meaning representation. *Brain and Language,* 86, 347–365.

Wu, L., & Barsalou, L. W. 2009. Perceptual simulation in conceptual combination: evidence from property generation. *Acta Psychologica*, 132 (2), 173–189.