

Clustering verbal Objects: manual and automatic procedures compared

Iliaria Colucci

University of Pavia

Department of Humanities

[colucci.ilaria03@gmail.com]

Elisabetta Jezek

University of Pavia

Department of Humanities

[jezek@unipv.it]

Vít Baisa

Lexical Computing Ltd.

Czech Republic

[vit.baisa@gmail.com]

Abstract

As highlighted by Pustejovsky (1995, 2002), the semantics of each verb is determined by the totality of its complementation patterns. Arguments play in fact a fundamental role in verb meaning and verbal polysemy, thanks to the sense co-composition principle between verb and argument. For this reason, clustering of lexical items filling the Object slot of a verb is believed to bring to surface relevant information about verbal meaning and the verb-Objects relation. The paper presents the results of an experiment comparing the automatic clustering of direct Objects operated by the agglomerative hierarchical algorithm of the Sketch Engine corpus tool with the manual clustering of direct Objects carried out in the T-PAS resource. Cluster analysis is here used to improve the semantic quality of automatic clusters against expert human intuition and as an investigation tool of phenomena intrinsic to semantic selection of verbs and the construction of verb senses in context.

Keywords: Clustering, verbal Objects, Italian, Semantic Types

1 Introduction

Clustering techniques have been used extensively in recent decades in Linguistics and NLP, especially in Word Sense related tasks. As a matter of fact, partitioning data sets on the basis of their similarity at a distributional level clarifies the meaning of lexical elements (Brown et al., 1991). Partitioning verbal arguments, for example, can be beneficial to investigate the sense properties they share but also to explore verbal meaning. In fact, as highlighted by Pustejovsky (1995, 2002), the semantics of each verb is determined by the totality of its complementation patterns and arguments play a fundamental role in verb meaning and verbal polysemy, thanks to the sense co-composition principle. *Id est*, the process of *bilateral semantic selection* between the verb and its

complement gives rise to a novel sense of the verb in each context of use (*ibidem*).

Clustering lexical items filling the argument positions of a verb is then believed to bring to surface relevant information about verbal meaning and the verb-arguments relation. Clustering them, and especially direct Objects in pro-drop languages such as Italian, allows hence to investigate how to better induce, discriminate and disambiguate verb senses. Because argument fillers share the same semantic nature, they can be grouped and generalized with respect to their content and be associated with semantic types, i.e. empirically identified semantic classes representing selectional properties and preferences of verbs.

Clustering of Objects can therefore be used as a survey tool for the intrinsic phenomena of semantic classes and, at the same time, as an object of investigation to improve the clustering automatic models themselves against human partitioning. This paper presents the results of an experiment comparing manual and automatic clustering of Italian Object fillers to be used in verb-sense identification and, along with it, it describes the linguistic phenomena that emerged from the semantic analysis of non-supervised clusters. The comparison concerns the agglomerative hierarchical clustering algorithm of the Sketch Engine corpus tool¹ (Kilgarriff et al., 2014) and the manual clustering carried out in the T-PAS resource² (Ježek et al., 2014), in which verbal senses are identified in context based on the fillers of the argument positions (see section 1.1) and are annotated with a semantic type (ST; see section 1.2) able to identify them. Thanks to their semantic generalization properties, ST are also believed to represent a useful comparative tool between manual and automatic clustering. After presenting the theoretical background of the research, section 2 will cover

¹ www.sketchengine.eu/

² tpas.fbk.eu

data, method and work pipeline, while clustering evaluation via metrics and linguistic analysis will be presented in section 3.

1.1 Clustering verbal Objects fillers

Clustering is a Data Mining task (Kotu & Deshpande, 2014) in which a grouping process of a set of objects is carried out, obtaining clusters of elements which are similar to each other but dissimilar from the objects of other groups (Xu & Wunsch, 2008). In most implementations, clustering is used with an exploratory function, i.e. it is a technique applied to data sets for which there is no *a priori* knowledge concerning the set membership of the samples (Lavine and Mirjankar, 2006). In these cases, clustering is therefore considered as a non-supervised procedure with the aim of providing an insight into the studied data. However, it can be considered a supervised method and regarded as a classification task when a manually created benchmark (a *ground truth*) is used to assess the output of the clustering (Bishop, 1995). The manually created partition or the manually defined set of classes is used to validate the groupings proposed by the automatic algorithm, through a process defined as external clustering evaluation (Gan, Ma and Wu, 2007). The idea behind this paper is to operate through a procedure very similar to external evaluation in which the manual clustering and the automatic one taken into consideration are mutually compared; but yet here the aim is not to validate the automatic model but more to bring out matches and differences between the partitioning criteria at the basis of the supervised clustering and the unsupervised one.

The supervised clustering under consideration here was performed on the lexical items that fill different argument positions in T-PAS, a resource of predicate-argument structures for Italian obtained from corpora (Ježek et al., 2014). T-PAS contains, for each argument slot, the specification of the semantic class to which the fillers found in that position in the corpus belong. We considered the direct Object clusters, which therefore contain the fillers that occupy that slot in the various occurrences of the corpus. To clarify this, given the following sense for verb *pilotare* (to pilot), the related cluster for the Object position will appear as follows:

(1) *pilotare*

1. [Human] *pilotare* ([Flying Vehicle] | [Water Vehicle])

pilotare_clust1: {macchina (car), moto (motor-bike), barca (boat), caccia (fighter aircraft), nave (ship)}

The ST defined for the direct Object slot can thus also be used as a label to semantically identify what is contained in the cluster.

As for automatic clustering, in our comparison we used the built-in clustering function (Baisa et al., 2015) in the Sketch Engine tool (SkE). The model is based on a hierarchical agglomerative algorithm that compute the distributional similarity³ between the Object fillers and groups them in an unsupervised way, starting from a minimum similarity value given to the algorithm (Kilgarriff et al., 2014). Clusters creation starts with computing Word Sketches, i.e. automatic, corpus-based summaries of a word's grammatical and collocational behaviour (Kilgarriff et al., 2004). The results concerning the direct Object are then grouped through a bottom-up process in which clusters are populated through pairings of words. The inclusion and exclusion criterion is a minimum default value of 0.15⁴ for distributional similarity. The clusters created in Sketch Engine for *pilotare* are the followings, for which, unlike T-PAS, ST labelling is not available:

(2) *pilotare_clust1*: {nave (ship), barca (boat)}
pilotare_clust2: {macchina (car), moto (motorbike)}
pilotare_clust3: {caccia (fighter aircraft)}⁵

The main difference between T-PAS and SkE clustering procedures are the semantic-distributional criteria on which they are based. T-PAS approach can be defined as verb-oriented: Objects are primarily clustered on the basis of their verbal distributional behaviour and ability to activate a given verbal sense as direct objects. Since all fillers occupying a given slot for a given sense share the same relation with the verb, they can be ontologically and semantically generalized with an ST on the basis of their common semantic traits. This generalization allows to make the verbal selectional constraints visible. On the contrary, SkE

³ See Kilgarriff et al. (2015) for statistics and technical details on similarity computing.

⁴ We also conducted a similarity value manipulation experiment, which confirmed what discussed in detail in section 3.

⁵ Clusters consist at least of 1 word and up to 1000.

performs noun-based clustering: it takes into account the general distributional behaviour of fillers, not merely the verbal one. In the process of creating sets, each filler behaviour is weighed against the entire reference corpus and with respect to the frequencies of appearance in different contexts. The elements clustered together in SkE are therefore not only similar in their sense and behaviour as direct objects, but also respect to the whole nominal class they belong to.

1.2 T-PAS System of Semantic Types

As mentioned above, in T-PAS argument slots are linked to ST labels, semantic classes able to generalize over the sets of lexical items in argument positions found in the corpus (Ježek et al., 2014). The labels belong to the System of Semantic Types (see Figure 1 for an excerpt), a hierarchical structure of semantic categories achieved by performing the CPA procedure (Hanks, 2004), on the evidence of 1200 Italian verbs (Ježek, 2019), i.e. through the manual analysis of examples in corpora of slots’s fillers and their co-occurrence statistics. They characterize a group of lexical elements with respect to their content, defining also a criterion of similarity and dissimilarity on which T-PAS clusters are created. STs are used here as a reference for the comparison of the two clustering models, for the verification of the clusters internal semantic quality.

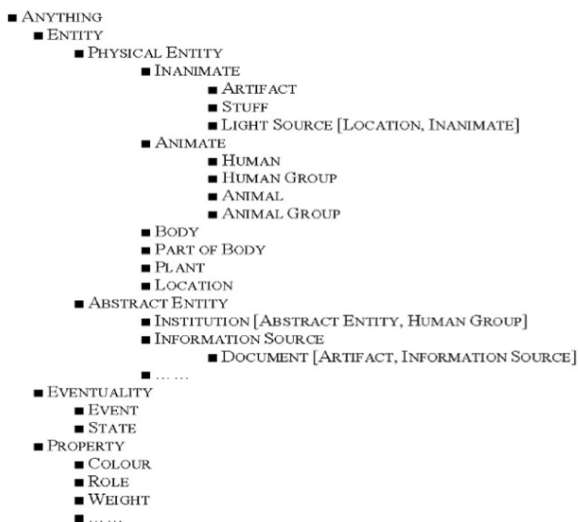


Figure 1: Top-level nodes and a selection of leafs from the ST System (Ježek, 2019)

2 Data and method

The research has been developed through a pipeline organized according to the following steps:

1. *Data extraction*: Data for both clusterings are extracted from the web crawled corpus ItWac reduced (Baroni et al., 2009). In this early stage the clusters of Object fillers for each verb included in T-PAS are extracted from the corpus annotated lines, while for Sketch Engine, the clusters are extracted for all verbs present in the ItWac corpus. All lines in the corpus are then scanned and verbal Objects are mapped through the condition: *OBJ = post verbal noun* (PostV_N). Since T-PAS does not annotate individual fillers as such but only works at verb and sentence level, this function is also used to retrieve its Objects.

2. *Data intersection*: The obtained clusters are intersected with each other in order to obtain a database in which there are sets for the same verbs and containing the same fillers, to focus on how the two models carried out the partition.

3. *Data filtering*: In this step the database is cleared from:

- verbs with structures recognized as complex and non-compositional, i.e. idiomatic constructions;
- verbs with the ST [Anything] (top node in Fig. 1) in the object slot, as it does not entail selection restrictions within the T-PAS clusters;
- verbs with Object clusters with more than 29 internal elements.

At the end of the filtering process the clusters of the two models are aligned with respect to the STs, i.e. all possible STs signaled in T-PAS for the Object of a verb are treated as a single set of semantic conditions, in order to analyze the internal quality of SkE clusters through them. The aligned structure of the verb *acquire* (to acquire) in Figure 2 is given here as an example.

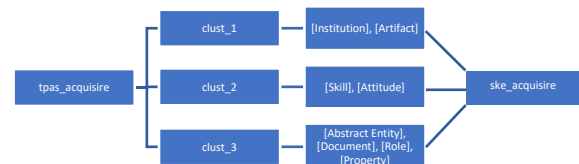


Figure 2: Aligned STs structure of verb *acquire*

The final database comes to a total of 397 verbs and 3938 clusters, including both T-PAS and SkE clusters. We provide an illustrative table (Table 1)

showing the first and last verb among those analyzed and the information on their respective clusters: *abbagliare* (to dazzle) and *votare* (to vote).

Verb	Source	Clusters	Clustered items	Items nr.
abbagliare	T-PAS	1	nemico, viso, utente	3
abbagliare	T-PAS	2	cliente, inquisitore, uomo, visitatore	4
abbagliare	SkE	occhio	nemico, uomo	2
abbagliare	SkE	inquisitore	inquisitore	1
abbagliare	SkE	pilota	visitatore, cliente, utente	3
abbagliare	SkE	viso	viso	1
votare	T-PAS	1	barzioletta, foto, poesia, sito	4
votare	T-PAS	2	riduzione, emendamento, legge	3
votare	SkE	barzioletta	barzioletta	1
votare	SkE	post	poesia, foto	2
votare	SkE	sito	sito	1
votare	SkE	mozione	emendamento	1
votare	SkE	risoluzione	legge	1
votare	SkE	bilancio	riduzione	1

Table 1: Complete set of clusters of the first and last verb of the data set

3 Clustering evaluation

3.1 SkE clustering evaluation

To verify the compatibility between the two clusterings, the similarity between the two partitions has been evaluated through different metrics able to offer an external evaluation of the unsupervised model. To account for both the presence of common pairings, as well as the homogeneity and completeness of the clustering, the following metrics were considered: Fowlkes & Mallow Index (F&M), Adjusted Rand Index (ARI), Homogeneity, Completeness.

F&M, as the geometric mean between precision and recall, was used to verify the similarity between the two models from how many partition pairings are in common. This index also allows to better balance the possible noise or unrelatedness between clustering (Fowlkes & Mallows, 1983). ARI (Hubert & Arabie, 1985) always gives information on the overlapping of the two clusterings in comparison but balances the very large number of clustered elements in T-PAS (Romano et al., 2016). Homogeneity and completeness metrics (Rosenberg & Hirschberg, 2007) are helpful to better investigate the internal content of the SkE clusters. They allow to highlight a possible internal structure, hierarchically and semantically coherent with the taxonomy identified for ST. Ho-

mogeneity evaluates if all automatic clusters created contain only elements that are members of a single class in the manual reference. Completeness, instead, evaluates if all the objects that are members of a given cluster in SkE are elements of the same cluster in T-PAS.

As reported by their respective creators, all metrics have an optimal result range between 0 and 1. The possible results between these two limits can be classified with respect to the greater or lesser proximity to the optimal limit: the results closer to 1 denote greater similarity of output between the two models, the results closer to 0 instead less similarity (Gan, Ma and Wu, 2007).

In this sense, we can define three bands of possibilities, coherently with the approach *the higher the better* generally used in cluster analysis: from 0.01 to 0.399, the clustering compared to the golden standard is highly different, from 0.4 to 0.699 the result and the correspondence is medium-good, while the results above 0.7 and up to 0.999 are the ideal ones, which indicate a marked correspondence between the compared models. However, since metrics such as F&W and ARI have shown the lack of partitions in higher ranges (the first beyond 0.82, the latter beyond 0.7), we choose to consider the whole group of medium good results between 0.4 and 1. The absence of the higher ranges stands for low compatibility between the two models.

Metric	Clusters in the [0.4-1] range
Adjusted Rand Score	11.08%
Fowlkes & Mallow	36.18%
Homogeneity	94.96%
Completeness	41.31%

Table 2: Metrics results

As we see in Table 2, what we find in fact is a situation of only limited correspondence between the two clustering, with a rather low overlap and similarity as indicated by the ARI and the F&M, even with the internal noise balance. At least two reasons may be behind the scarce similarity: the tendency of SkE to create small fine-grained clusters populated by few elements that give more weight to specificity than to generalization capacity; the fact that in T-PAS for a given verb sense the Object slot can be compatible with more STs (see (1)), and such STs can also be hierarchically distant in the general system of labels. This leads to clusters containing fillers able to activate a

given verbal sense but which are quite heterogeneous among themselves and semantically dissimilar, with respect to the rest of the distributional relations between the fillers. An example can be the verb *trasportare* (to transport), which has as T-PAS cluster for the first sense a set of 18 Objects (see (4)); such fillers belong to three different STs: [Inanimate], [Animate] and [Energy]. The latter ST, [Energy], is hierarchically distant to the others since it has a different parent node than [Animate] and [Inanimate], which both pertain to a lower level in the hierarchy.

(3) *trasportare*

1. [Human] | [Vehicle] | [Watercourse] *trasportare* [Inanimate] | [Animate] | [Energy]

(4) *trasportare_clust1*: {acqua (water), alimento (nourishment), animale (animal), arma (weapon), bene (asset/good), bicicletta (bicycle), cadavere (corpse), cibo (food), gas (gas), gommone (inflatable raft), macchina (machine), oggetto (object), peso (weight), student (student), terra (soil), traffico (traffic), viaggiatore (traveller), visitatore (visitor)}

It is clear that a fine-grained algorithm, not able to generalize at a higher level as in T-PAS, will divide fillers labelled with [Animate] or [Inanimate] from those labelled with [Energy]. In fact, SkE for the same verb creates 12 clusters.

As shown in Table 2, the results of the Completeness are in line with what has just been discussed for ARI and F&M: only in 40% of the cases all members of a T-PAS cluster are members of a single SkE cluster. These are generally small or medium sized clusters with only one associated ST or with hierarchically close alternative ST structures. Homogeneity highlights the primary characteristic of SkE clusters and the algorithm: it is preferable to create smaller but internally purer clusters, rather than larger sets with members of other classes. This implies the creation in SkE of semantically specific clusters, that privilege the inter-relation between Object fillers but not the higher semantic level between Object fillers and verb.

From a different perspective, we can say that the noun-oriented criteria of clustering and the verb-oriented ones tend to converge when we consider small clusters, in which the elements belonging to a set in SkE generally belong to the same set in T-PAS.

As for wide clusters, they are particularly rare in SkE and tend to be smaller in size than T-PAS anyway. Their content also seems to be dependent on various factors on which the linguistic analysis has shed light.

3.2 Linguistic analysis of the clusters

To verify the nature of the diversity between the two clusterings measured with the metrics reported in 3.1, a detailed analysis of the lexical-semantic phenomena visible internally to the clusters was carried out considering:

- The consistency, for automatic clusters, with one and only one of the aligned T-PAS STs, i.e. the precision and purity at the semantic level of clusters compared to the generalization of the ST;
 - Internal homogeneity, i.e. whether the clusters meet verb-sense oriented or noun-sense oriented criteria and, if the latter, whether the cluster items are linked by syntagmatic relationships and there is some kind of affinity or implication between them. Thus, the types of semantic relations present between the words are identified;
 - The overlap between clusters with respect to the ARI, and in relation to cluster size and clustering difficulty depending on several STs possible for the same slot;
 - The problem of incorrect mapping as Objects of postverbal Subjects, subjects of inaccusative verbs, structures with *si* particle (e.g. reflexive, impersonal), i.e. the clusters' internal noise.
- The research has shown that SkE clusters tend to be small-medium sized, semantically homogeneous, often able to isolate very specific semantic relations. They are generally not consistent *per se* with the verb sense identified by T-PAS but create partitions: a) usually of medium size and consistent with only one parallel ST, b) single element groups that generally belong to a higher level of specificity or to a different semantic domain, and c) groups that are inconsistent with the sense of the verb but cluster words on the basis of the following criteria:
- Belonging to the same domain (e.g. informatics for *distribuire* {software, applicazione});
 - Being part of the same ST, but as very specific instances, not separated by the T-PAS hierarchy (e.g. {abbazia, monastero, santuario} for *saccheggiare* and the type [Location]);
 - The possibility of a conceptual association or affinity (e.g. {seminario, incontro, seduta} for *organizzare*);

- Purely distributional parameters and undefined semantic relations (e.g. in *gestire* {contenuto, caso});

- A relationship of synonymy or meronymy (e.g. {spinta, propensione} for *frenare* or for *fratturare* {dito, mano, braccio}); antonymy, hyponymy, hyponymy are generally represented by different clusters.

The parameters of consistency, internal homogeneity and overlapping between the models seem to relate to the same factors: first, the size of the clusters, i.e. how many clustered elements are part of the set; second, the structure of STs possible for the Object (see (5)), i.e. if for the same slot only one ST is possible, if several alternatives are available or, also, if a lexical set is signaled in the T-PAS annotation - that is, if among the fillers a set of lexical elements is present that has high frequency or has the typical behaviour of a collocation (e.g. {messaggio | ricordo} in (5)). This is relevant since the computation of SkE starts precisely from the frequency and collocational behaviour of a word.

(5) *cancellare* (sense description: to eliminate, to make inexistent):

1. [Human] | [Inanimate1] | [Abstract Entity1] | [Eventuality1] *cancellare* [Inanimate2] | [Abstract Entity2] {messaggio | ricordo} | [Eventuality2]

The third relevant factor is hierarchical proximity, i.e. if STs possible for a slot are sisters of the same parent node between the types present in the hierarchy (see (6)).

(6) no proximity: [Animate] vs. [Institution]
in proximity: [Command] vs [Request]

The clusters of SkE, even if not corresponding to those of T-PAS, are rated totally consistent or at 70-80% consistent with one of the aligned STs 62% of the times; in the remaining 38% of cases, there is a significant number of clusters that can be generalized with a ST. Consistency is more difficult to reach if a given sense is annotated in T-PAS with several alternative STs or STs and lexical sets co-presenting; SkE can produce new combinations in which fillers corresponding to different STs are included in the same cluster. Very frequently SkE atomizes the set of fillers in nuclear clusters, made up of only one or two elements that are necessarily consistent with one ST but are not of much help to the study of semantic relations.

As said, hierarchical proximity of STs and the size of the cluster can influence its handling: if the possible STs for the Obj-slot are hierarchically distant and the T-PAS cluster is small, the SkE outcome will tend to be heterogeneous and inconsistent. Consider the verb *affogare* (to drown), which has [Animate] and [Emotion] as possible STs for the same slot but different senses. T-PAS clusters are small, *clust_1* counts 3 elements and *clust_2* counts 7, in which the two possible types are clearly distinguished. One would expect to find the two senses separated in the SkE clusters as well, since [Emotion] belongs to lower levels of the hierarchy and has a different parent node of [Animate]. However, for SkE we find clusters such as:

(7) *affogare_clust3*: {figlio (son), bimbo (child), pensiero (thought)}

If, on the contrary, we consider closer ST, the clustering will be homogeneous, even if not verb-sense-oriented because too fine-grained.

Medium-sized clusters (more than 10 clustered elements) seem to perform quite well both with hierarchically close and distant STs. A useful example can be *smarrire* (to lose) in T-PAS (8), for which SkE presents the clusters in (9):

(8) *smarrire*

1. [Human] *smarrire* [Artifact]
2. [Human] | [Human group] *smarrire* [Concept] | [Property]

(9) *smarrire_clust1*: {borsello (man bag)}

smarrire_clust2: {significato (meaning), ragione (reason), memoria (memory), pensiero (thought)}

smarrire_clust3: {capacità (capacity), consapevolezza (awareness)}

smarrire_clust4: {nozione (notion), certezza (certainty)}

smarrire_clust5: {senno (sense)}

smarrire_clust6: {fiducia (trust), voglia (will)}

smarrire_clust7: {documento (document)}

smarrire_clust8: {cellulare (mobile phone)}

Considering that for *smarrire* T-PAS creates only two clusters (see (10)), the sets in (9) also highlight the general tendency of SkE to create smaller, semantically highly fine-grained clusters.

(10) *smarrire_clust1*: {borsello (man bag), cellulare (mobile phone), documento (document)}

smarrire_clust2: {consapevolezza (awareness), capacità (capacity), certezza (certainty), fiducia (trust), memoria (memory), nozione (notion), pensiero (thought), ragione (reason), senno (sense), significato (meaning), voglia (will)}

Large clusters are generally difficult to handle because ideally portionable in more distributionally cohesive groups. As regards the problem of internal noise, due to the *PostV_N* relation, we can note that the phenomenon is pervasive and important, since it affects the results of internal coherence and homogeneity. It is, however, a phenomenon that can be curbed with a revision of the extraction function. What emerges from the analysis is a distance in the general structure of the two clustering results but a good compatibility from the internal semantic point of view. T-PAS privileges rather more complex semantic groupings on a level of co-composition between verb and meaning, linked to conceptual operations of generalization. On the contrary, SkE creates complex and homogeneous structures of relations inside data, even if sometimes this implies clusters that are too fragmented and not always optimal also from a noun-oriented perspective. T-PAS seems to pertain to a higher level of granularity respect to SkE, whose clusters can be considered as possible sub-partitions of the STs.

4 Conclusion

The paper presented the statistical and linguistic results of a comparison between SkE unsupervised clustering model and the manual and verb-sense oriented clustering of T-PAS. It highlighted how the noun-oriented model and the verb-oriented one are not overlapping if not partially. The SkE clustering, even if not overlapping, can still be considered as internally compatible with the T-PAS partition, since the homogeneity metric reaches good results. The internal linguistic analysis allowed to identify the semantic quality through the consistency with a semantic type, the internal homogeneity, the adherence with the verb-oriented approach of T-PAS. The reasons that regulate the fragmentation of clusters in SkE, i.e. motivations that follow a fine-grained logic, were then presented. The analysis made possible to shed a light on the semantic compatibility between the two approaches, which seem to pertain to different levels of granularity.

The difference in the partition output and the parallel semantic compatibility allows us to claim

that the SkE automatic clustering is more useful for the internal investigation of STs than to investigate the verb-Object co-composition relation. It would be interesting to conduct further comparisons between other automatic clustering techniques and that of T-PAS, to investigate additional semantic implications of clustering through noun-based and verb-based approaches.

Reference

- Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2009). The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. In *Language resources and evaluation*, 43(3):209-226.
- Baisa, V., El Maarouf, I., Rychlý, P., & Rambousek, A. (2015). Software and Data for Corpus Pattern Analysis. In *RASLAN*, 75-86.
- Bishop, C. M. (1995). *Neural networks for pattern recognition*. Cambridge UK. Oxford University Press.
- Brown, P., Della Pietra, S., Della Pietra, V., and Mercer, R. (1991). Word sense disambiguation using statistical methods. In *Proceedings of the 29th Meeting of the Association for Computational Linguistics (ACL-91)*, 264-270.
- Fowlkes, E. B., & Mallows, C. L. (1983). A method for comparing two hierarchical clusterings. In *Journal of the American statistical association*, 78(383):553-569.
- Gan, G., Ma, C., & Wu, J. (2007). *Data clustering: theory, algorithms, and applications*. Society for Industrial and Applied Mathematics.
- Hanks, P. (1996). Contextual dependency and lexical sets. In *International journal of corpus linguistics*, 1(19):75-98.
- Hanks, P. (2004). Corpus pattern analysis. In *Euralex Proceedings*, 1:87-98.
- Hubert, L., & Arabie, P. (1985). Comparing partitions. In *Journal of classification*, 2(1):193-218.
- Ježek, E. (2019). Sweetening Ontologies Cont'd: Aligning bottom-up with top-down ontologies. In *JOWO*.
- Ježek, E., Magnini, B., Feltracco, A., Bianchini, A., & Popescu, O. (2014). T-PAS: A resource of corpus-derived Types Predicate-Argument Structures for linguistic analysis and semantic processing. In *Proceedings of LREC*, 890-895.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kořář, V., Michelfeit, J., Rychlý, P., Suchomel, V. (2014). The Sketch Engine: ten years on. In *Lexicography*, 1:7-36.

- Kilgarriff, A., Baisa, V., Bušta, J., Jakubiček, M., Kovář, V., Michelfeit, J., Rychlý, P., Suchomel, V. (2015). *Statistics used in Sketch Engine*. (document available at: <https://www.sketchengine.eu/wp-content/uploads/ske-statistics.pdf>)
- Kilgarriff, A., Rychly, P., Smrz, P., and Tugwell, D. (2004). The Sketch Engine. In *Information Technology*, 105:116-126.
- Lavine, B. K., & Mirjankar, N. (2006). Clustering and classification of analytical data. *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation*.
- Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge MA. MIT Press.
- Pustejovsky, J. (2002). Syntagmatic processes. In Cruse, A. D., Hundsnurscher, F., Job, M., Lutzeier, P. (eds.) *Lexicology: A Handbook on the Nature and Structure of Words and Vocabularies*. de Gruyter.
- Romano, S., Vinh, N. X., Bailey, J., & Verspoor, K. (2016). Adjusting for chance clustering comparison measures. In *The Journal of Machine Learning Research*, 17(1):4635-4666.
- Rosenberg, A., & Hirschberg, J. (2007). V-measure: A conditional entropy-based external cluster evaluation measure. In *Proceedings of EMNLP-CoNLL 2007*, 410-420.
- Xu, R., & Wunsch, D. (2008). *Clustering*. Hoboken NJ. John Wiley & Sons.