Author Proof

With or Without Meaning? Hype Cycles in Language Technology and What We Can Learn from Them

António Branco(✉)

Department of Informatics, Faculty of Sciences,
University of Lisbon, Lisbon, Portugal
Antonio.Branco@di.fc.ul.pt

AQ1

Abstract. Despite its relatively short period of existence as a scientific area, natural language processing has gone through a succession of diverse mainstream research paradigms. How similar are these inflection moments in the history of the research on language technology? What can we learn from that similarity, if any, about the overall shape of the evolution of this field? And importantly, what can we anticipate from this shape, if any, about the future and emerging trends in language technology? — which is the topic of the workshop where this paper was presented.

The result of this study is meant to be of help to organize research agendas of centers, laboratories and individual researchers and innovators, as well as to guide informed institutional funding and support for research and innovation in language technology. AQ2

Keywords: Hype cycles · Scientific progress · Future and emerging trends · Natural language processing · Language technology · Computational linguistics

1 Introduction

The Association for Computational Linguistics (ACL) is a leading organization of researchers and professionals natural language processing, a field whose object of research is the computational mapping between linguistic form and meaning and the language processing applications that such mapping can support. This association organizes an annual meeting in a different country across the globe every year. An important moment of these conferences is the plenary session with the presidential address, in which the president typically shares his views about the mission of the association and the emerging challenges and opportunities for this area.

In contrast to the conference programs of previous meetings, the program for the 2015 conference, which was held in July in Beijing, looked like having been flooded by papers resorting to, supported or inspired by deep learning techniques. The president of ACL at the time, Chris Manning, Professor at

J.F. Quesada et al. (Eds.): FETLT 2016, LNAI 10341, pp. 1–17, 2017.
https://doi.org/10.1007/978-3-319-69365-1_1

Stanford University and a leading researcher in the area, though he had been exploring deep learning for language technology in his own work for some time, seems to have been impressed by what this might represent as a tectonic change in the direction that the field might be taking — so much so that he devoted his presidential address (Fig. 1) to share what he understood that was and could be the relation between natural language processing (NLP) and connectionist techniques.

In a nutshell, he argued that, given the nature of natural language, the application of deep learning techniques to its processing should not be expected to lead to the high level of gains that were obtained in their application to other domains, such as vision or speech processing, where the state of the art performance more than doubled. His talk might have been motivated by the greatly increased weight of deep learning papers in that year's conference program, but it was also seen as a reaction to the big wave of immoderate optimism which was both being motivated and ridden by the promise of deep learning for the unlimited progress of any domain where it might be applied to.

Fig. 1. Picture of the President of ACL in the opening address at the 2015 ACL meeting, Beijing.

In that year of 2015, a few months before that conference in Beijing, in one of its May issues, the Nature journal included a dissemination paper with the title "Deep Learning" by leading researchers in that area [1]. Defending the superiority of (unsupervised) deep learning, they stated that:

> "Human learning is largely unsupervised: we discover the structure of the world by observing it, not by being told the name of every object."

While it may be indeed not correct that humans discover the structure of the world just by being told the name of every object, that sensible acknowledgment by itself, however, does not make it correct, in turn, that humans discover the structure of the world just by observing it — as centuries of research on the theory of knowledge and human learning have helped us to understand. But rather than being taken as a key contribution for the theory of human learning, this statement helps to illustrate how optimism and a certain ideological reductionism may go hand in hand, and emerge even in the ranks of skillful scientists.

Interestingly, ideological reductionism backing the concentration of research efforts into a given technique or methodology — represented by the above quotation —, and eliciting exercises of devising upper bounds for the success of that technique by its mere analytical inspection — illustrated by the plenary talk referred to above — induces a déjà vu feeling. Though with a different methodology and players, the same type of jolt was experienced in the area of language technology (LT) around two decades ago, with the advent of statistical approaches to natural language processing.

This apparent analogy, together with the questions that it elicits, is driving the analytical exercise of the present paper: How similar are these two moments in the history of the research on language technology? What can we learn from that similarity, if any, about the evolution shape of this field? And importantly, what can we anticipate from this shape, if any, about the future and emerging trends in language technology? — which is the theme of the workshop where the invited plenary talk corresponding to this paper was presented.

In the next Sect. 2, we elaborate on what may be the basic elements that can be found as driving the evolution of research on language technology.

Section 3 will be devoted to represent the effect of those drivers along a timeline covering the history of this field, which will support the exercise of identifying hype cycles and their respective triggers.

In Sect. 4, we ponder on the attractors and deflectors of these hype cycles in language technology, which are supported by the research on cognition at large, while in the following Sect. 5, we ponder in turn on the enablers of scalable language technology solutions, which are supported by innovation in information technology in general.

On the basis of the materials and evidence collected in the sections preceding it, in Sect. 6 we discuss what we consider to be the emerging enabler and the ultimate attractor of language technology, which may help to devise the direction of future trends in this field.

In the last Sect. 7, we close the paper and its prospective study with an indication of what, in our view, follows from the analysis undertaken along the paper as the emerging trend for the area of natural language processing.

The outcome of this prospective analysis is meant to be of help to organize research agendas of centers, laboratories and individual researchers and innovators, as well as to guide informed institutional funding and support for research and innovation in language technology.

2 Swinging Back and Forth Between Form and Meaning

Let us start with the first question put forward above.

The advent of statistical approaches (statistical NLP) and later on, the advent of connectionist approaches (neural NLP) are two salient moments of accelerated change in the research on natural language processing. As in many respects they are inducing a somewhat déjà vu feeling among practitioners who participated in both, what can be motivating that feeling? How similar are these two moments in what they represent to the history of the research on language technology?

As made evident by the quotation above in Sect. 1, the perceived key advantage of neural NLP is to be able to do without supervised training. As supervised training relies critically on linguistically interpreted and annotated data sets (e.g. syntactic treebanks, semantically annotated corpora, etc.) — which approximates in different degrees a representation of the meaning conveyed —, the perceived supreme advantage of neural NLP is thus to rely only on input raw linguistic forms to be fully operative, thus dispensing with linguistic analysis and ultimately with the need of prepared linguistic features and specifically designed linguistic and semantic representations.

Dispensing with linguistic analysis, and a fortiori with a representation of meaning, was also the promise that prompted the enthusiastic optimism around statistical NLP in its initial days. At those times, the research work that had been carried out on meaning representation and processing was termed as symbolic NLP. That was the "old school" with respect to which statistical NLP was initially seen as bringing superior advantages by focusing on raw linguistic forms, and thus by allowing for a more streamlined and directly accessible research field, liberated from the cumbersome intermediation of linguistic generalizations and meaning representations.

Analogously, neural NLP is exercising its initial attraction by distancing itself from distributional NLP with its focus on designing and obtaining the representations of the meaning of natural language words and expressions.

In both occasions there seems to be a movement of swinging away from the representation of meaning, and towards focusing on raw linguistic forms. Of course, the repetition of this movement was possible because in-between there has been a pendular change, from statistical NLP to distributional NLP, where the representation of meaning regained its momentum again (Fig. 2).

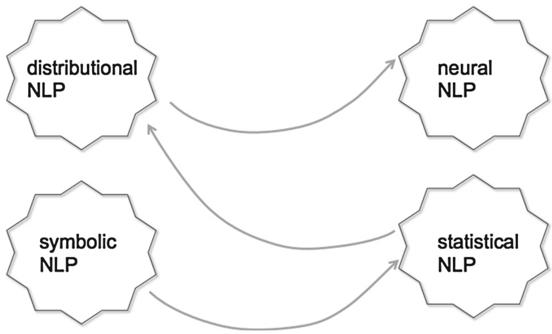


Fig. 2. Sequence of mainstream NLP paradigms.

Given these considerations, we should turn to the second question put forward in the previous section: What can we learn from these similarities about the shape of the evolution of the language technology field?

Breaking free? One possible interpretation could be that this circumstance indicates that this field is advancing by successive superior paradigms that successively replace previous, inferior ones.

But that is not the only possible analysis.

Encircled? A less positive view is that this putting into perspective of different moments of the research on language technology allows to bring to light that this field is actually not making substantial progress after all, as it may be stagnated with competing paradigms that oppose each other, with no essential advancement.

Some publications could be brought in support of this view. For instance, in their paper on "Improving Distributional Similarity with Lessons from Word Embeddings", Levy et al. report on an exercise of systematic and controlled comparison of statistical and connectionist approaches to similarity under distributional semantics [2], concluding that:

> "... we observe mostly local or insignificant performance differences between the methods, with **no global advantage to any single approach** over the others"

Spiraling forward? Yet another view could be that the language technology area is actually advancing, but by new paradigms extending previous ones rather than by replacing them. An example of a paper that could support this view is "A Study on Similarity and Relatedness Using Distributional and WordNet-based approaches", whose experiments by Agirre et al. [3] indicate:

> "that distributional similarities can perform as well as the knowledge-based approaches, and a **combination of the two can exceed the performance of results previously reported** on the same datasets"

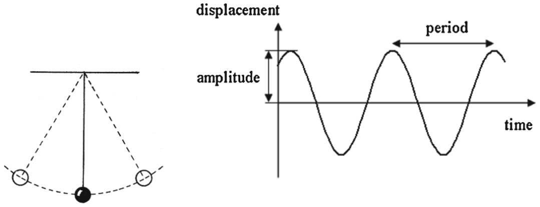


Fig. 3. Representations of oscillatory motion. (Credits: https://en.wikibooks.org)

As it often happens in science, changing the representation of a state of affairs may be crucial for a novel insight into its key ingredients and thus for gaining a more thorough understanding of it. In what concerns for instance an object in physical oscillatory motion, this movement can be captured by representing the points where the object may be along its trajectory (Fig. 3, left), or by bringing also time into the representation and thus by recording its displacement with respect to the central position along the time in a two-axis graphic (Fig. 3, right).

As noted above, language technology has evolved in a pendular fashion, with stronger foci either on the representations and processing of linguistic meaning or on the processing of linguistic forms. Taking inspiration from the figures above, it may be serendipitous to depict those transitions along a timeline represented in the x-axis, with the y-axis representing the predominant emphasis of the approach, either on the side of form or on the side of linguistic meaning (Fig. 4).

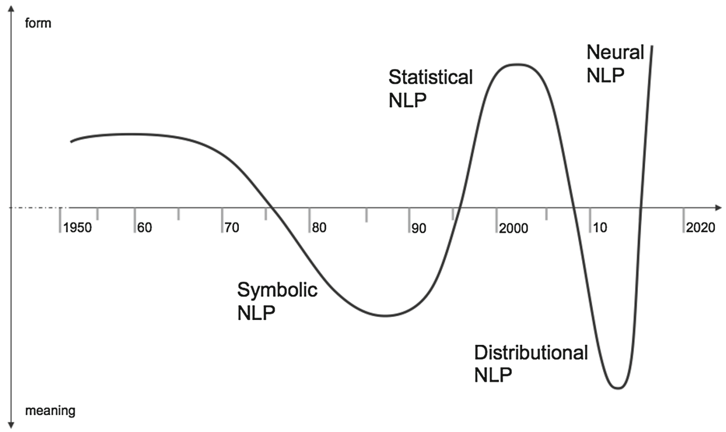


Fig. 4. Hype cycles in language technology.

There has been more than one phase in the research on language technology where the mainstream focus was on the meaning or on the form. That is rendered by there being more than one peak both above and below the x-axis.

Also, a more recent peak on a given side of the x-axis has a higher amplitude than the previous peak on the same side thus rendering that more recent approaches to NLP based on meaning, respectively on form, are more sophisticated and explore more intensively the meaning, respectively the form, relations.

Additionally, more recent peaks have shorter wavelengths than previous ones. This reflects the increasing frequency of the advent of novel research approaches in the field of language technology.

3 Hype Cycles in LT and Their Emblematic MT Triggers

Our third driving question is about what can be anticipated from this shape concerning the future and emerging trends in language technology. In order to get a better vantage point to address this issue, it is worth having first a more informed understanding of the context and triggers of the hype cycles underlying that shape.

It is common wisdom among practitioners of language technology that machine translation (MT) is a most demanding application in natural language processing, deemed to be its quintessential challenge as it virtually calls for the articulation of a whole range of partial results and processing tasks in this domain. Very interestingly, key inflection points shaping the hype cycles in language technology can be traced back to key papers that introduced novel MT paradigms (Fig. 5).

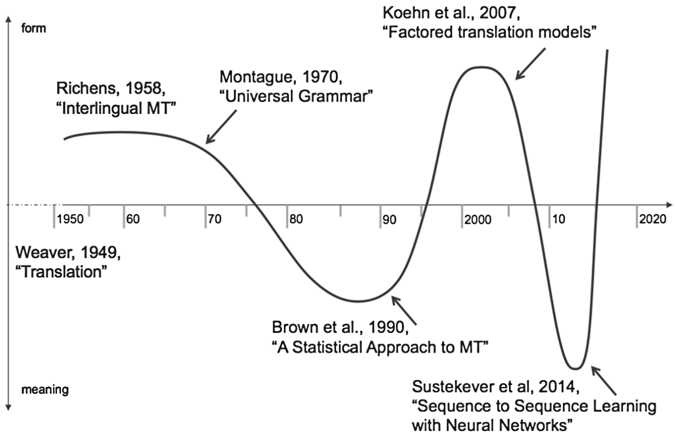


Fig. 5. Hype cycles in language technology and their emblematic machine translation triggers.

3.1 Meaning Transfer

The proposal towards an MT whose crucial step is based on the transfer of meaning representations dates back at least to the four page seminal paper of Richens from 1958, "Interlingual Machine Translation" [4]. But it would be the article "Universal Grammar" from 1970 by Montague [5] that may be better seen as one of the emblematic triggers of a consistent trend towards meaning oriented MT, and away from previous approaches based on some version of word-into-word replacement.

For any sentence of a fragment of English, this paper showed how it was possible to algorithmically obtain its translation into a logical language. As the resulting logical formulas allow to model key semantic relations, these were thus primary candidates of representations of meaning that are independent of particular natural languages. Hence, this represented a major encouragement for the exploration of the MT model based on the transfer of meaning, or at least in some abstract enough representation of its linguistic properties along the so-called Vauquois triangle (Fig. 6).

$$\langle h0, \{h1\!:\!\mathrm{every}(x, h2, h3), h4\!:\!\mathrm{dog}(x), h5\!:\!\mathrm{probably}(h6), h7\!:\!\mathrm{chase}(x, y),$$
$$h8\!: \mathrm{some}(y, h9, h10), h11\!: \mathrm{white}(y), h11\!: \mathrm{cat}(y)\},$$
$$\{h0 =_q h5, h2 =_q h4, h6 =_q h7, h9 =_q h11\}\rangle$$

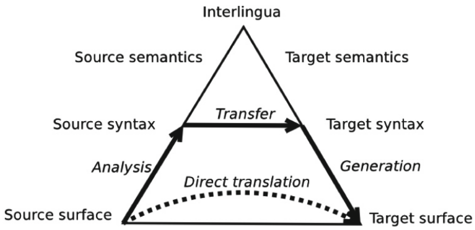


Fig. 6. Top: Meaning representation of the example sentence Every dog probably chases some white cat in the Minimal Recursion Semantics description formalism [6], p. 302. Bottom: Vauquois triangle. (Credits: http://mttalks.ufal.ms.mff.cuni.cz/images/f/f1/Pyramid.png)

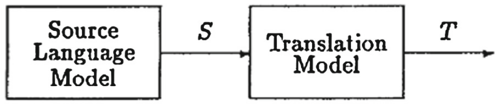
To a large extent, this was also the pervasive model in symbolic NLP. The key assumption is that for any substantive application or problem in language technology — from summarization to conversational interfaces —, it should be addressed ideally by resorting to the intermediation of some explicit representation of the meaning of its input.

3.2 Co-occurrence Inferencing

MT envisaged as a possible instance of some stochastic model may be traced back at least to 1949, when Weaver wrote his memorandum titled "Translation" [7]. But it would be the article by Brown et al., "A Statistical Approach to MT", from 1990 [8], which would become an emblematic inflection point and set in motion a consistent and increasing interest in exploring MT under this paradigm that focuses on the linguistic form and moves away from the previous emphasis on the representation of meaning.

Under this paradigm, the noisy-channel is the basic underlying model, which is used in speech recognition and that had its origin in Shannon's work in 1948 about correcting errors in the communication of messages [9]. The motivating goal is to recover a string that got distorted as a consequence of its transmission through a noisy communication channel: This recovering is undertaken on the basis of the combination of stochastic models of the language to which the string belongs (supporting the prediction of what string components follow each other) and of the communication channel (supporting the prediction of what string components might have been erased, inserted or replaced) (Fig. 7, top).

In an MT set up, the language model concerns the language targeted by the translation process and the channel model concerns the translation model based on possible replacements, and respective probabilities, between the source language and the target language words (Fig. 7, bottom).



$$\Pr(S) \quad \times \quad \Pr(T \mid S) = \Pr(S,T)$$

das		**Haus**		**ist**		**klein**	
e	*t(e|f)*	*e*	*t(e|f)*	*e*	*t(e|f)*	*e*	*t(e|f)*
the	0.7	*house*	0.8	*is*	0.8	*small*	0.4
that	0.15	*building*	0.16	*'s*	0.16	*little*	0.4
which	0.075	*home*	0.02	*exists*	0.02	*short*	0.1
who	0.05	*household*	0.015	*has*	0.015	*minor*	0.06
this	0.025	*shell*	0.005	*are*	0.005	*petty*	0.04

Fig. 7. Top: Diagram of noisy-channel model applied to machine translation, from [8] p. 80. Bottom: Examples of possible lexical translation probabilities in a translation model, from [10] p. 84.

This paradigm shift in MT inspired the application of stochastic inferencing techniques to language technology, which eventually induced the advent of statistical NLP. Ultimately relying on frequencies of co-occurrences of linguistic forms in collections of utterances, this approach brings the emphasis to handling forms and their surface quantitative relations in detriment of language processing intermediated by some degree of meaning representation.

3.3 Linguistic Factors

The mainstream approach to MT and NLP eventually found an inflection point with the inclusion into the stochastic models of quantitative information on more linguistic features and generalizations that were increasingly more abstracted away from raw linguistic forms. This hybridization of statistical and symbolic approaches eventually encompassed the representation of meaning.

The processing of meaning and its possible representation gained thus a revival, this time adding a new angle to it, namely under the perspective of distributional semantics (Fig. 8 (b)), where the meaning of expressions and their semantic relations are based on high dimension vectors ultimately relying on frequencies of some co-occurrences.

As in previous changes of focus in mainstream language technology research, this inflection point may be traced back to some emblematic trigger related to MT, like the paper in 2007 by Koehn et al., "Factored Translation Models" [12]. The log-linear model for MT proposed in this publication (Fig. 8 (a)) introduced a new paradigm for MT, where the impact of linguistic factors can be integrated into the translation procedure, including highly abstract linguistic information on underlying syntactic and semantic features.

a)

$$p(x) = \exp \sum_{i=1}^{n} \lambda_i h_i(x)$$

b)

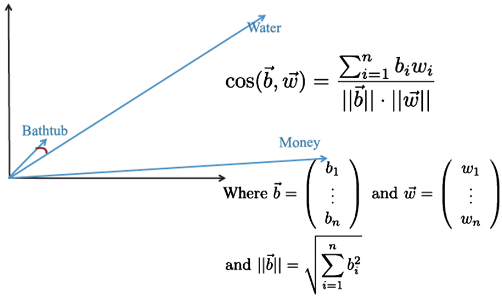


Fig. 8. (a) Log-linear model structure, where x is the random variable, λi are weights, and hi feature functions that can be instantiated with language and translation models as well as with models for relevant "linguistic factors", from [10] p. 138. (b) Computing similarity between vectors representing the meaning of example words as the cosine of the angle between them, from [11] p. 636.

3.4 Encoding-Decoding

As the mainstream direction of research on natural language processing became oriented towards resorting to some form of intermediary meaning representation, the oscillatory pattern that seems to underlie the research in this field became again apparent. A turning point redirected once again the focus of interest, this time towards approaches based on linguistic forms and their mere surface relations, and dispensing with specifically designed representations of linguistic meaning.

Such inflection eventually emerged and once again a key contribution came from the MT area, in this case by an emblematic paper in 2014 by Sustekever et al., "Sequence to Sequence Learning with Neural Networks" [13], whose title emphasizes the focus on the modeling of the relations between the forms (source and target "sequences"), circumventing the need of handling some linguistic representation.

For each expression of a sequence of input to be translated, the recurrent neural network computes an internal interim output that takes into account that expression and the previous interim output, where in the last step, concerning the last expression in the input sequence, a word of the output sequence is also output. From that point onwards, the model computes the next step by taking the last internal interim output and last expression emitted into account. The sequence of expressions emitted constitutes the proposed translation of the initial input sequence (Fig. 9).

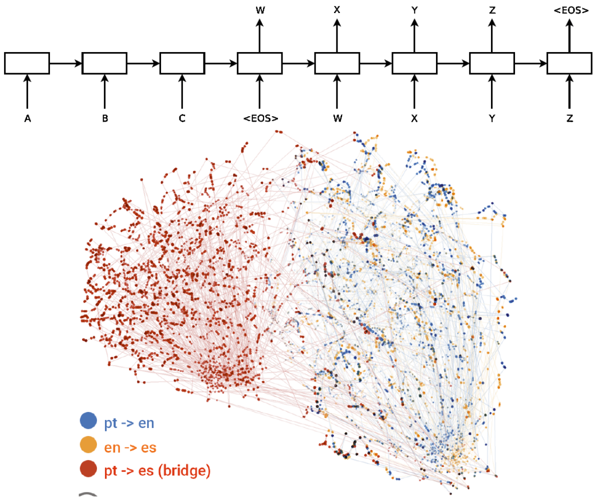


Fig. 9. Top: Schematic diagram of a translation model as a recurrent neural network, where A-C stand for expressions in sequence of the input sentence in the source language, and W-Z are expressions of the output sentence in the target language, from [13] p. 3105. Bottom: Two-dimension projection of high-dimensional vectors where each point represents a single decoding step during the translation process and where points that represent steps for a given sentence are connected by line segments, from [14] p. 10.

4 Cognition at Large: LT Attractors and Deflectors

Following the considerations above, it is enticing to associate the evolution of the research direction of language technology to emblematic publications on MT. Since MT is considered a quintessential NLP application, it is only natural that new paradigms for MT would set the example for new kinds of approaches, and thus be important drivers of change, for the whole field of language technology.

While there may be such a triggering or driving effect by new paradigms for MT, it is nevertheless also worth noting that these novelties are themselves triggered or enhanced by much broader underlying changes that occur outside the language technology (LT) area proper, namely in the broader and encompassing area of cognition technology at large. Accordingly, in as much as the LT inflection points may be fostered by new MT paradigms, like what happens with the emergence of the latter, such LT inflection points appear also strongly influenced by broader changes and successes in cognition technology at large (Fig. 10).

In this connection, it is worth noting that the advances and successes in neighboring areas, like speech or vision processing, have certainly acted as important attractors helping in the inflection of the direction of mainstream language technology.

The success of statistical methods in speech processing stimulated research seeking similar success in applying such methods first to MT and then in a generalized way to the whole field of language technology. The inflection towards

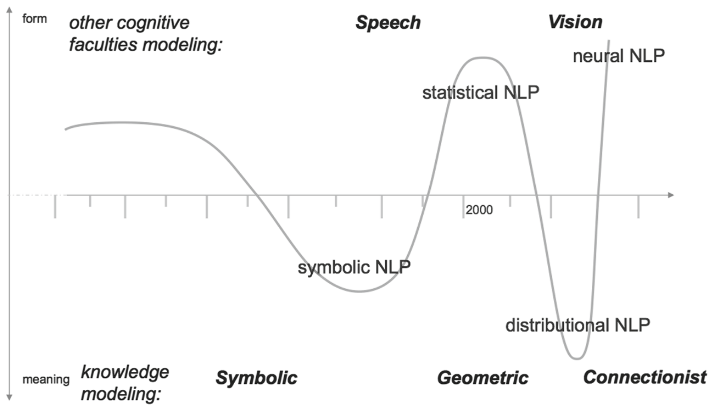


Fig. 10. Hype cycles in language technology and attractors from other areas in cognition technology.

statistical NLP in the 1990's was greatly influenced by that attempt of emulating the success achieved in speech processing with statistical approaches.

Some two decades later, in the 2010's, further advances in speech processing, where accuracy was doubled with the application of a neural networks approach, would act again as an important attractor, now towards neural NLP. This time that effect was compounded with concomitant success in vision processing, as also in this area the advent of deep learning permitted important advances.

Interestingly, while the modeling of other neighboring cognitive faculties — speech and vision — and its success pushed the focus of NLP towards linguistic form, knowledge modeling and the advent of its diverse paradigms, in turn, acted as attractors towards mainstream NLP based more on linguistic meaning.

Until the 1990's, when the focus in language technology started being displaced to linguistic form, the mainstream approach to knowledge modeling was based on symbolic methods and this was concomitant with the emergence of symbolic NLP.

When the pendular oscillation brought again linguistic meaning into central focus in the 2000's, with distributional NLP, the geometric approach to knowledge modeling might have played also an important inspirational role.

5 Information Technology at Large: Enablers of Scalable LT

As language is a core cognitive faculty and language technology is a central area in the realm of cognitive technology, it is thus natural that the advances and successes of research in neighboring cognitive faculties and respective research areas — such as speech and vision processing and knowledge modeling — and the methodological innovations underpinning them have influenced the direction of language technology research, and thus played the role of attractors, or deflectors, in what emerges as its pendular shape of development, leaning to giving primacy, in alternation, either to form or to meaning.

While there certainly is such an influence by new approaches from other cognitive technologies, it is nevertheless also worth noting that these novelties are themselves enabled by much broader underlying changes that occur outside the cognitive technology realm proper (Fig. 11).

In this connection, it is worth noting that for the consolidation of statistical NLP started in the 1990's, the advent and generalization of the internet played a crucial role as this permitted the accumulation and availability of increasingly larger and richer language datasets, without which statistical approaches for language technology could not have matured.

By the same token, it is the advent and generalization of computational devices with increasingly large storage capacity and more processing speed that enabled the viability of distributional NLP, since the 2000's, and currently of the emergence of neural NLP, which are both based on data and time intensive computational procedures.

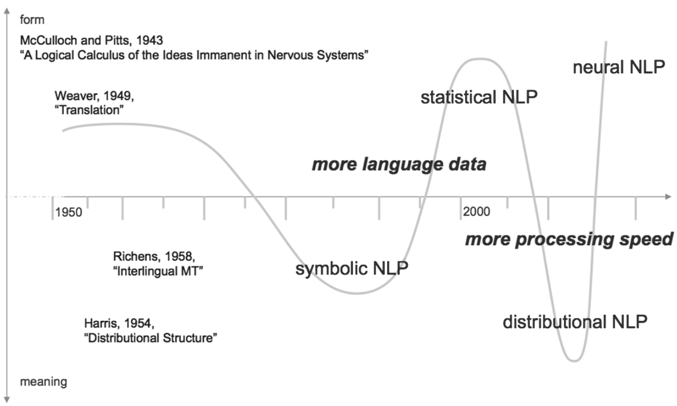


Fig. 11. Hype cycles in language technology and enablers from information technology for seminal proposals of different NLP paradigms.

That these developments in Information Technology at large have had a crucial enabling role for scalable language technology is as more salient as the key methodological insights for distributional and neural NLP had been published long time before, namely more than half a century before.

The key insight for distributional NLP can be traced back at least to 1954, to the paper "Distributional Structure" by Harris [15]. And in the case of neural NLP, as for the whole connectionist endeavor, the seminal ideas from 1943 by McCulloch and Pitts, published in the paper "A Logical Calculus of the Ideas Immanent in Nervous Systems" [16], are a landmark.

6 Emerging Enabler and Ultimate Attractor

As the objective of the present paper is to contribute to the reflection on the future and emerging trends in language technology, the discussion expanded in the previous sections concerns the analysis of previous development in language technology and is thus instrumental to address this objective.

The previous sections helped to make evident the oscillatory shape, between form and meaning, of the focus of the research on natural language processing, and that such shape is influenced by enabling factors from information technology and by attraction factors from cognitive technology. Accordingly, in order to pursue our objective, it is worth pondering on the possible forthcoming enablers and attractors.

In our view, the strongest candidate to be the **forthcoming key enabler** for language technology, contributed from information technology at large, is the advent of the internet of things (Fig. 12).

The advent of the internet helped to accumulate ever larger amounts of linguistic data. This has enhanced the processing of the relations among linguistic expressions and permitted to evolve from **compositional** semantics, the mainstream approach to linguistic meaning until the 2000's, to **distributional** compositional semantics, which emerged in the 2010's.

Likewise, the advent of the internet of things will help to accumulate ever larger amounts of data about extra-linguistic objects, including about their individual proper names (e.g. id numbers, IP addresses, nicknames assigned by their owners or users, etc.), their features (e.g. their color, shape, age, use, etc.) and about the relations among them (e.g. their location, their proximity to each other and to the speakers, their previous mentioning together, etc.). Expectedly, this will enhance the processing of the relations between linguistic expressions and their referents, in all their challenging forms, including deictic reference, contextualized definite descriptions, anaphoric relations, etc. Accordingly, this will permit to evolve from the current mainstream distributional compositional semantics, to a novel approach for the analysis and processing of the meaning of natural language, namely to **referential** distributional compositional semantics.

As for the **forthcoming key attractor** for language technology, contributed from cognitive technology at large, in our view this will result from the ongoing efforts and eventual achievements in the area of knowledge modeling of coming up with **unified cognitive models**, which will amplify the strengths and mitigate the drawbacks of the symbolic, geometric and connectionist contributions.

The impact of the internet of things, as the forthcoming information technology enabler, and of the unified cognitive models, as the forthcoming attractor from cognitive technology, will be convergent in bending the direction of the development of language technology back to be centered around the representation and processing of meaning.

7 LT: Spiraling Forward with Cross Hybridization

The considerations expanded above in Sects. 2 to 5 concern the analysis of previous development in language technology and are instrumental in reflecting about the future trends in language technology, the central objective of this paper. In the previous section, those considerations were instrumental in making evident the recurrent influence of enablers and attractors that are **external** to language technology, and thus in looking for the eventual emergence of forthcoming enablers and attractors and their anticipated impact.

Likewise, those considerations will be instrumental also in making evident the modulation that is **intrinsic** to language technology in terms of the evolving direction of the research in this area, which should also be explored to ponder on its future trend.

Shorter periods. Mainstream approaches to natural language processing have alternated between a stronger focus either on the form or on the meaning of linguistic expressions. As can be more easily observed from its graphical representation (Fig. 12), the pace of this oscillation has not been constant. The changes of paradigm have been succeeding at a faster pace along the decades, with these oscillations happening at an increasingly shorter period. It is likely this trend continues into the future.

Wider amplitude. Language technology has been driven more than once by a stronger focus on form, respectively on meaning. Interestingly, a revival of a stronger focus on form, respectively on meaning, does not mean a mere return to previously explored solutions but a leap forward with denser relations in form, respectively in meaning, being captured and explored. This trend receives a graphical representation by means of oscillations with wider amplitude and should also continue into the future.

Increased integration. The curve in the plot is an abstraction deemed to represent the direction of mainstream natural language processing. While at any

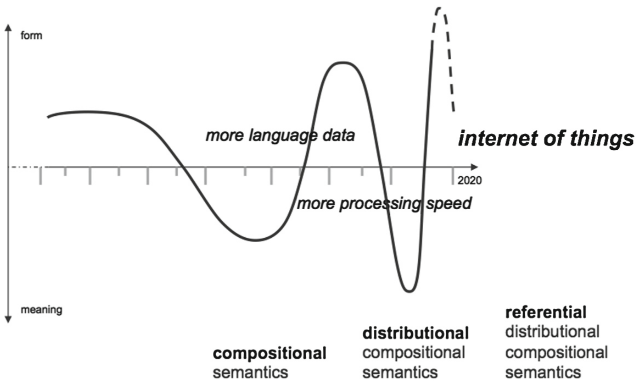


Fig. 12. Hype cycles in language technology, emerging enabler and future trend.

time interval the mainstream focus of interest is represented as moving away from a certain type of approach (and towards another type of approach), it is not the case that the previous results happen to be abandoned altogether. The inflection movements are associated with inherent latencies, which lead to temporal overlaps of different foci and approaches, and importantly, which allow for their hybridization.

Hence, increasingly shorter periods and wider amplitude in the evolution shape of language technology indicate a trend, to be intensified into the future, of an increasing and accelerating integration between the diverse aspects related to the representation and processing of form and meaning, and an increasing and accelerating cross-fertilization among the diverse paradigms.

This can be illustrated by a few recent emblematic examples of hybridizations in MT. For instance, the paper by Devlin et al. in 2014 is an exercise on combining statistical and neural approaches [17]; in 2016, Dong and Lapata, in turn, pave the way for articulating transfer and neural solutions [18]; and Gaudio et al. in 2016 report on advances on hybridization of transfer and statistical based methods [19].

More powerful semantics. The above discussion and analysis helps to understand that the hype cycles in language technology — oscillating between form and meaning — are likely the manifestation of an underlying long term trend of spiraling forward with cross hybridization of approaches and results that are advancing the representation and processing of the relation between linguistic form and meaning, which is the ultimate cornerstone of natural language processing.

Taking into account this intrinsic longstanding evolutive shape, its current direction of progress, and the anticipated impact of forthcoming external enablers and attractors, the future trend in language technology would likely direct its development to be based on a deeper meaning representation more densely anchored in linguistic form.

This more powerful semantics should result from the hybridization between sentential and compositional semantics, evolving from symbolic NLP, lexical and conceptual semantics, contributed by distributional NLP, and future referential and situated semantics, supported by neural NLP.

Better on-the-go applications. The performance of diverse types of NLP applications should benefit from this more powerful semantics.

It is likely that some of them eventually enjoy remarkable progress or even gain a new twist, especially those whose functionality is more related or dependent on a situated usage. That will be likely the case of conversational interfaces, in particular in on-the-go environments and with autonomous agents, be they artificial or human, especially in multilingual settings.

Acknowledgments. The work reported in this paper was partially supported by P2020 Program under the grant 08/SI/2015/3279 for the project ASSET-Intelligent Assistance for Everyone Everywhere.

References

1. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**, 436–444 (2015)
2. Levy, O., Goldberg, Y., Dagan, I.: Improving distributional similarity with lessons learned from word embeddings. Trans. Assoc. Comput. Linguist. **3**, 211–225 (2015)
3. Agirre, E., Alfonseca, E., Hall, K., Kravalova, J., Pas, M., Soroa, A.: A study on similarity and relatedness using distributional and WordNet-based approaches. In: Proceedings of Human Language Technologies: 2009 Annual Conference of the North American Chapter of the ACL, Boulder, pp. 19–27, June 2009
4. Richens, R.: Interlingual machine translation. Comput. J. **1**(3), 144–147 (1958)
5. Montague, R.: Universal grammar. Theoria **36**(3), 373–398 (1970)
6. Copestake, A., Flickinger, D., Pollard, C., Sag, I.: Minimal recursion: an introduction. J. Res. Lang. Comput. **3**(4), 281–332 (2005)
7. Weaver, W.: The Rockefeller Foundation (1949). Translation AQ3
8. Brown, P., Cocke, J., Pietra, S.D., Pietra, V.D., Jelinek, F., Lafferty, J., Mercer, R., Roossin, P.: A statistical approach to MT. Comput. Linguist. **16**(2), 79–85 (1990)
9. Shannon, C.: A mathematical theory of communication. Bell Syst. Tech. J. **27**(3), 379–423 (1948)
10. Koehn, P.: Statistical Machine Translation. Cambridge University Press, Cambridge (2010)
11. Erk, K.: Vector space models of word meaning and phrase meaning: a survey. Lang. Linguist. Compass **6**(10), 635–653 (2012)
12. Koehn, P., Hong, H.: Factored translation models. In: Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Prague, pp. 868–876, June 2007
13. Sustekever, I., Vinyals, O., Lee, Q.: Sequence to sequence learning with neural networks. In: Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS 2014), Montreal, 8–13 December 2014, pp. 3104–3112 (2014)
14. Johnson, M., Schuster, M., Le, Q., Krikun, M., Yonghui, W., Chen, Z., Thorat, N.: Google's multilingual neural machine translation system: Enabling zero-shot translation (2016). arXiv:1611.04558v1
15. Harris, Z.: Distributional structure. Word **10**(2–3), 146–162 (1954)
16. McCulloch, W., Pitts, W.: A logical calculus of the ideas immanent in nervous systems. Bull. Math. Biophys. **5**, 115–133 (1943)
17. Devlin, J., Zbib, R., Huang, Z., Lamar, T., Schwartz, R., Makhoul, J.: Fast and robust neural network joint models for statistical machine translation. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, 23–25 June 2014, pp. 1370–1380 (2014)
18. Dong, L., Lapata, M.: Language to logical form with neural attention. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, 7–12 August 2016, pp. 33–43 (2016)
19. Del Gaudio, R., Labaka, G., Agirre, E., Osenova, P., Simov, K., Popel, M., Oele, D., van Noord, G., Gomes, L., Rodrigues, J.A., Neale, S., Silva, J., Querido, A., Rendeiro, N., Branco, A.: SMT and hybrid systems of the QTLeap project in the WMT16 IT-task. In: Proceedings of the ACL 2016 First Conference on Machine Translation (WMT 2016), Association for Computational Linguistics, Berlin, 11–12 August 2016, pp. 435–441 (2016)