

A Corpus-based Analysis of “Create” and “Produce”

Siaw-Fong Chung*

Abstract

This paper examines the synonymous pair ‘create’ and ‘produce’ in English and suggests that their similarities and differences can be elucidated based on the types of products denoted by their objects. PRODUCTS, as part of the eventive information of ACTIVITY, are sometimes shared by both verbs (e.g., in ‘Manufacture’), while some other products (e.g., ‘Farming’) are found only in one of the verbs (‘produce’). In addition, a gradience analysis of ‘property,’ ‘creativity,’ ‘quantity,’ and ‘concreteness’ is also postulated, as these semantic features help to decide the selection between ‘create’ and ‘produce.’ Two corpora – the Brown Corpus and the Frown Corpus (from ICAME) – provided the data for this paper, and their results were compared to the British National Corpus through the collocation application of the Sketch Engine. Finally, the similarities and differences between the compared words are represented by means of a hierarchical structure.

Keywords: Near-synonym, Corpus, Create, Produce, Sense Analysis, Gradience Analysis

* Assistant Professor, Department of English, National Chengchi University, E-mail: sfchung@nccu.edu.tw. The author wishes to express her thanks to two anonymous referees for their invaluable comments on this paper. Special gratitude also goes to Kathleen Ahrens for her comments on the previous version of this work. The research herein is supported partly by NSC grants project number 100-2628-H-004-137-, 99-2410-H-004-206-, and 97-2410-H-004 -001-.

1. Introduction

Synonyms are usually identified by linguists and lexicographers through substitution. Other methods include the use of dictionary meanings in distinguishing senses. However, neither substitution nor comparison of dictionary meanings is able to provide quantitative information in a set of data such as distributional information of object types, argument structures, and frequencies of linguistic patterns. When using substitution, one possible problem is the non-existence of “perfect synonyms” (term used in Taylor, 2002: 262). “Perfect synonyms,” or “absolute synonyms” in Lyons’ (1995: 61) terminology, refer to a pair of synonyms in which (a) “all meanings [being compared] are identical”; (b) two words are “synonymous in all contexts”; and (c) they are “semantically equivalent (i.e., their meaning or meanings are identical) on all dimensions of meaning, descriptive and non-descriptive.” The presence of all these characteristics, as also noted by Taylor (1995: 56), is rare: “[P]erfect synonyms – lexical items with the same meaning and which are therefore interchangeable in all contexts – are exceedingly rare.”

Most of the synonyms are likely to be “near” synonyms rather than “perfect” synonyms. Near-synonyms, as Cruse (1986: 267) defines them, are “lexical items whose senses are identical in respect of ‘central’ semantic traits, but differ [...] in ‘minor’ or ‘peripheral traits’.” Therefore, there will be “central” shared meanings between near-synonyms, in addition to the differing “peripheral” meanings. For this present work, it is our aim to identify these central and peripheral meanings, as well as to display them using a hierarchical representation, through a gradience analysis of semantics.

In near-synonym research, traditional approach usually involves discrete semantic features such as [\pm human] and [\pm animacy]. However, in some cases, it was found that dichotomies as such might not solve the near-synonym problem. This is the case in our analysis of ‘create’ and ‘produce’ in which a continuum-based concept must be applied. The two research questions asked in this research are as follows:

- (1) (a) What are the similarities and differences between the verbs ‘create’ and ‘produce’?
- (b) How can we represent the results from a corpus-based methodology that differentiates the meanings of ‘create’ and ‘produce’?

These two research questions will be answered based on quantitative linguistic analyses utilizing corpora data.

2. Approaches to Distinguishing Near-Synonyms

Near-synonyms can be examined through descriptive comparison and quantitative analysis. Collinson (1939, mentioned in Harris, 1973: 14) was an example of earlier descriptive studies, which used semantic features to distinguish synonyms. Collinson’s list consisted of elements such as ‘general/specific applicability,’ ‘intensity,’ ‘emotion,’ ‘moral approbation,’ ‘professionalism,’ ‘written/non-written,’ ‘colloquialism,’ ‘local/dialect,’ and ‘child talk.’ Some of these (e.g., ‘general/specific’ and ‘written/non-written’) are still commonly used, while others (e.g., ‘intensity,’ ‘emotion,’ and ‘colloquialism’) are more often discussed at discourse level. Features such as ‘moral approbation’ and ‘child talk’ are currently seldom used as semantic features.

Unlike descriptive research, quantitative research on near-synonyms usually involves computation of collocations or experimental results. Xiao and McEnery (2006), for example, used collocation and semantic prosody to examine near-synonyms. Biber et al. (1998), too, had examined the use of ‘begin’ and ‘start’ using the Longman-Lancaster Corpus. A computational-based approach can also be seen in Church et al. (1994) in which the verbs ‘request’ and ‘ask for’ were compared in terms of substitutability. In addition, there are also studies such as Taylor (2002), which carried out experiments in order to test similarities and differences of near-synonyms. Taylor contrasted the adjectives ‘tall’ and ‘high’ using an acceptability rating task in which subjects were asked to rate whether the use of these two adjectives was acceptable under different contexts. Taylor later argued that the differences between these two verbs can be captured using MacLaury’s (1997, 2002) Vantage theory. Using this theory, Taylor claimed that the dominant (“fixed landmark” in both) and the

recessive (“dimensional uses” for ‘tall’ and “positional uses” for ‘high’) meanings of the two adjectives could be found.

Whether the approach is descriptive or quantitative, the ultimate aim of near-synonym research is usually to differentiate closely related meanings. Cruse (1986: 271-277) stated that meanings can be differentiated by comparing (a) the interface between speakers’ intentions and languages; and (b) the interactions among linguistic items (detailed in (2) below).

- (2) (a) Interface between speakers’ intentions and languages
 - (i) Propositional mode (which depends on types of statements, such as question, simple statement, exclamation, command, etc.)
 - (ii) Expressive mode (which conveys emotion, attitude, register effect, or formality style)
- (b) Interactions among linguistic items
 - (i) Presupposed meanings (which contribute to selectional and collocational restrictions)
 - (ii) Evoked meanings (which contribute to discourse cohesion and communicative roles)

This current work concentrates on the level of ‘presupposed meanings’ ((2b)(i) above) in which ‘selectional and collocational restrictions’ of the linguistic meanings between ‘create’ and ‘produce’ will be inspected. The approach adopted by our present paper is quantitative by first considering verb forms, followed by sense frequencies, and, finally, collocational analyses. We also compared the two verbs in terms of similarities by analyzing their sense frequencies in WordNet (version 3.0; cf. Fellbaum, 1998). We showed that, unlike the seemingly separated senses, the senses of each verb can be inter-related. We also tried to provide a list of semantic features necessary in distinguishing the two verbs. However, unlike discrete categories, these features are presented in continuums.

3. Methodology and Results

All instances used in this work were taken from corpus data. All verb forms of ‘create’ and ‘produce’ were searched using the AntConc concordance (Anthony, 2005) in two American corpora – the Brown Corpus (compiled in the 1970s) and the Freiburg-Brown (Frown) Corpus (documents of the early 1990s). These two corpora were comparable, as the Frown Corpus was designed to match the Brown Corpus. Information on both corpora can be obtained through the ICAME website (<http://icame.uib.no/newcd.htm>). From both corpora, 466 instances of ‘create’ and 481 instances of ‘produce’ were found. These instances were analyzed manually.

3.1 Verb Forms

The Brown and Frown corpora were searched using AntConc in order to retrieve all verb forms of ‘create’ and ‘produce.’ The following Table 1 provides the total number of instances collected from both corpora. (Sixteen instances of ‘produce’ as nouns, eight instances of ‘producing’ as adjectives, and two instances of ‘created’ as adjectives were removed manually before the results in Table 1 were analyzed.) The number of instances found in each corpus is presented and percentages are given in brackets.

From Table 1, one can see that both corpora are consistent in terms of verb forms, with quite identical distributional patterns. The most often appearing forms are the *-ed* form (with 39.70% for ‘create’ and 41.16% for ‘produce’) and the bare infinitive form (with about 35% for both verbs). These two forms constitute about 75% of the total verb forms in respective corpus. Examples of each of these forms are given in (3) below. The other forms (*-s* and *-ing*) appear no more than 15% for both verbs.

- (3) (a) *Harvey said his objective was to **create** a better public image for welfare.*
(b) *The American League expansion **created**, inevitably, weaker teams.*
(c) *This method of rinsing appears to **produce** maximum cleansing with minimum soil redeposition.*
(d) *Then Quayle dropped in a paragraph that **produced** the spectacularly silly media effect...*

Table 1: Instances of ‘Create’ and ‘Produce’ from Brown and Frown Corpora

	‘Create’		
Verb Forms	Brown (%)	Frown (%)	Total (%)
Bare Infinitive form	53 (30.11)	111 (38.28)	164 (35.19)
-s form	13 (7.39)	35 (12.07)	48 (10.30)
-ed form	81 (46.02)	104 (35.86)	185 (39.70)
-ing form	29 (16.48)	40 (13.79)	69 (14.81)
Total	176 (100.00)	290 (100.00)	466 (100.00)
	‘Produce’		
Verb Forms	Brown (%)	Frown (%)	Total (%)
Bare Infinitive form	72 (33.96)	98 (36.43)	170 (35.34)
-s form	19 (8.96)	29 (10.78)	48 (9.98)
-ed form	90 (42.45)	108 (40.15)	198 (41.16)
-ing form	31 (14.62)	34 (12.64)	65 (13.51)
Total	212 (100.00)	269 (100.00)	481 (100.00)

In (3a) and (3c), examples for the infinitive form are given (although the figures in Table 1 also include the present tense of both verbs used with plural subjects). The purposive ‘to’ used in (3a) and (3c) denotes a process that is not yet completely executed; therefore, planning is undertaken so that a certain result can be expected through the process of creation/production. In (3b) and (3d), the past tense examples are given. For both verbs, the *-ed* form is often seen because the two verbs usually emphasize the results rather than the process. The use of this form, as we will introduce later, does not occur at random but, rather, is for specific purposes. Such motivation for using a grammatical form can be seen, among many, in Heine (1997) and Dixon (1991), who discussed the issue in further detail.

In addition to the verb forms presented above, we also found several similarities and differences in constructions. For both ‘create’ and ‘produce,’ they seem to allow the fronted-object constructions as demonstrated in (4) below.

- (4) (a) *to see what kind of a god these people will **create** or what strong convictions they will develop...*

- (b) *but the noise they **produce** can be kept to a minimum...*
- (c) *with the situation he had helped **to create**...*
- (d) *the handful of painters that Austria has **produced** in the 20th century...*

In all the sentences in (4), the objects are underlined. In some uses of prepositions following the two verbs, some common patterns are found as shown in (5) below.

- (5) (a) *Gaskell **created** such a world **for** herself*
- (b) *when Grayson **produced** a hat **for** Ryder*
- (c) *to **create** **for** a large national purpose*
- (d) *if you are going to **produce** **for** home use only*

The use of ‘for’ in (5a) and (5b) is benefactive, while the ‘for’ in (5c) and (5d) indicate purposes. Both examples in (5c) and (5d) are used intransitively, with the demotion of the objects being created and produced.

From our analysis, we found that some constructions are found only with one of the verbs. An example is given in (6) below, with the use of ‘out of’ for ‘create’ only in (6a). Instances of ‘produce’ with ‘out of’ were not found in the corpora. Finally, the use of ‘over-produce’ in (6b) is not seen with ‘create’ (*‘over-create’). In (6c), ‘upon’ is found only with ‘produce.’

- (6) (a) *our consumer society **created** an ethos of the self **out of** a need for ever more selfish consumers*
- (b) *the American farmer’s capacity to **over-produce** has cost the taxpayers a large amount*
- (c) *as a painting **produces** **upon** the gazer*

We also found the following middle-voice example in (7) below.

- (7) *the land cannot continue **to produce** with agribusiness practice*

Here, ‘the land’ is the subject that yields products. At the same time, ‘the land’ itself is also the goal of the cultivating process. Such uses are not seen in ‘create’ in our data (e.g., ‘*The drawing paper continues to **create** pictures/shades’).

In the sections that follow, results based on sense analyses and object types will be reported.

3.2 Frequencies of Senses

One of the advantages of comparing meanings between any two words is that it enables one to see whether there is an overlapping sense between the two words of interest. The existence of an overlapped sense may indicate that the two words are a near-synonymous pair. Our working hypothesis, therefore, is that there will be at least one sense that overlaps between these two verbs to make them a near-synonymous pair. (See also Chung and Ahrens (2008) for a similar claim on overlapping sense.) In order to prove this hypothesis, we first analyzed all instances of ‘create’ and ‘produce’ according to their different senses without referring to a dictionary. After the categories of meanings were produced, we then contrasted the categories with a dictionary source in order to label each sense. We selected WordNet 3.0, an online lexical resource that enables a search for words that are semantically related (cf. Fellbaum, 1998), and compared the meanings of both verbs with the meanings we found, respectively. The senses from WordNet are shown in (8) below.

(8) Senses from WordNet 3.0

(a) ‘Create’

- (i) **Make or cause to be or to become:** “make a mess in one’s office”; “create a furor”
- (ii) **Bring into existence:** “The company was created 25 years ago”; “He created a new movement in painting”
- (iii) **Pursue a creative activity; be engaged in a creative activity:** “Don’t disturb him—he is creating”
- (iv) **Invest with a new title, office, or rank:** “create one a peer”
- (v) **Create by artistic means:** “create a poem”; “Schoenberg created twelve-tone music”; “Picasso created Cubism”; “Auden made verses”

- (vi) **Create or manufacture a man-made product:** “We produce more cars than we can sell”; “The company has been making toys for two centuries”

- (b) ‘Produce’
 - (i) **Bring forth or yield:** “The tree would not produce fruit”
 - (ii) **Create or manufacture a man-made product:** “We produce more cars than we can sell”; “The company has been making toys for two centuries”
 - (iii) **Cause to happen, occur, or exist:** “This procedure produces a curious effect”; “The new law gave rise to many complaints”; “These chemicals produce a noxious vapor”; “The new President must bring about a change in the healthcare system”
 - (iv) **Bring out for display:** “The proud father produced many pictures of his baby”; “The accused brought forth a letter in court that he claimed exonerated him”
 - (v) **Cultivate by growing, often involving improvements by means of agricultural techniques:** “The Bordeaux region produces great red wines”; “They produce good ham in Parma”; “We grow wheat here”; “We raise hogs here”
 - (vi) **Bring onto the market or release:** “produce a movie”; “bring out a book”; “produce a new play”
 - (vii) **Come to have or undergo a change of (physical features and attributes):** “He grew a beard”; “The patient developed abdominal pains”; “I have strange spots all over my body”; “well-developed breasts”

As shown in (8a) and (8b) above, the overlapped senses (shaded) are ‘bring into existence/cause to happen, occur, or exist’ and ‘create or manufacture a man-made product.’ Based on the instances that we extracted from the corpora, we analyzed our data according to these senses. The results are shown in Table 2 below.¹

¹ Note that sense frequencies are also available in WordNet based on the Brown Corpus and Stephen Crane’s novella entitled *The Red Badge of Courage* (cf. Landes et al., 1998), but the

Table 2: Sense Distributions of ‘Create’ and ‘Produce’

Create				Produce			
Senses (8a)	WordNet 3.0	Freq.	%	Senses (8b)	WordNet 3.0	Freq.	%
i	Make/cause to be/to become	37	7.94	i	Bring forth or yield	48	9.98
ii	Bring into existence	294	63.09	ii	Manufacture	43	8.94
iii	Pursue a creative activity	14	3.00	iii	Cause to exist	271	56.34
iv	Invest with new title	29	6.22	iv	Bring out for display	27	5.61
v	Create by artistic means	88	18.88	v	Release to market	62	12.89
vi	Manufacture	4	0.86	vi	Cultivate	29	6.03
				vii	Grow/develop	1	0.21
	Total	466	100.00		Total	481	100.00

From Table 2, one can see that the most frequent sense is ‘bring into existence’ for ‘create’ (63.09%) and ‘cause to exist’ for ‘produce’ (56.34%). These most frequent senses constitute more than 50% of the total instances for each verb, respectively. Examples of these most frequently occurring senses are given in (9) below (with alternating verb added).

- (9) (a) *the particles of negative electricity, which by their action **create**[/**produce**] the forces that tie this atom of calcium to the neighboring atoms of oxygen...*
- (b) *a vision handicap that may **produce**[/**create**] nervous tension...*

analysis in this work also included the Frown Corpus. The use of WordNet’s sense categories in Table 2 is a result of the author not finding a significant difference between her own categories when compared to those from the WordNet, meaning the WordNet senses for both ‘create’ and ‘produce’ were found to be adequate in reflecting the categories created by the author. Nonetheless, the author provides additional distributional data to the existing WordNet senses.

In (9a) and (9b), both ‘create’ and ‘produce’ can be substituted by one another. As indicated by the philosopher Leibniz (1704, cited in Church et al. 1994: 154), “two things are identical if one can be substituted for the other without affecting the truth.” In this case, when the two verbs overlap, this proves that ‘create’ and ‘produce’ form a synonymous pair. This step is important because most previous studies either took synonymy directly from a dictionary or formulated a shared meaning intuitively. In those studies, the differences were usually the focus without first proving the proportion of the overlapped similarities. Our analysis, thus, added distributional data to WordNet’s existing senses.

In order to see further how collocations may become a factor in distinguishing the two verbs, the following section shows the analyses of object types. From the analyses, we found that many of the senses in Table 2, including the overlapping and non-overlapping meanings, could in fact be linked to one another through the concept of PRODUCT. (Small capitals are used for words denoting an event or components of an event [cf. Dowty (1979, 1991) and Parsons (1990)]. In a later section, we provide a possible solution to distinguishing ‘create’ and ‘produce.’ Our approach incorporates both collocation information and a gradience analysis of ‘create’ and ‘produce.’

3.3 Types of Objects

This section reports the analysis of the types of objects following the two verbs. Our analyses show that ‘create’ and ‘produce’ can be distinguished in terms of the concreteness of their objects. The results are shown in Table 3 below.

Table 3: Comparisons of Object Types of ‘Create’ and ‘Produce’

<i>Create</i>		<i>Produce</i>	
Objects	Freq. (%)	Objects	Freq. (%)
Abstract Objects	334 (71.52)	Abstract Objects	231 (49.36)
Concrete Objects	133 (28.48)	Concrete Objects	237 (50.64)
Total Objects	467 (100.00)	Total Objects	468 (100.00)
Total Corpora Instances	466	Total Corpora Instances	481

The total number of objects may differ from the total number of instances, given in Table 1, because coordinated objects such as ‘the value of goods and services produced’ were counted twice. However, these coordinated objects constitute less than 5% of the total number of objects for ‘create’ and ‘produce,’ respectively, and should not affect the overall patterns. The results in Table 3 also ignored the omission of objects such as in (10) below. From the closer number of total objects to total corpora instances in ‘create’ (467 and 466) compared to those in ‘produce’ (468 and 481), it is possible that the objects of ‘produce’ will be omitted more often (as in (10b)) than is the case for the objects of ‘create’ (e.g., (10a)).

(10) (a) *God **creates***

(b) *if you are going to **produce** for home use only...*

In examples such as in (10), the objects are unmentioned either because they are unimportant or because they are not the focus of the discussion. Many times, what is being ‘created’ or ‘produced’ can be inferred from the context and thus it is a pragmatic concern. Regarding object types, from Table 3, one can see that about 72% of the total instances of ‘create’ are followed by abstract objects (‘jobs,’ ‘world,’ ‘problem,’ etc.) while only 49% of the total instances of ‘produce’ are followed by abstract objects (‘effect,’ ‘results,’ ‘chang,’ etc.). Examples of types of objects are provided in (11) below.

(11) (a) *Uncle Sam would then accuse them of **creating** a monopoly by “unfair competition” (abstract)*

(b) *worthy of a pastry chef **creating** a wedding cake for the marriage of [someone] and [someone] (concrete)*

(c) *and regulatory changes are **producing** significant effects, according to the study (abstract)*

(d) *a new generation of farmers who will **produce** diversified crops that are in demand (concrete)*

From Table 3, it seems clear that the dichotomy of concrete and non-concrete (or

[±concrete]) cannot be applied here, as both are present in the verbs. As Langacker (1994: 9) had defined, “[a] continuous parameter has the property that, between any two values (however close), an intermediate value can always be found.” It is different from “discreteness,” which “implies a direct ‘jump’ between two distinctive values, one of which is nonetheless the immediate successor of the other.” This paper, thus, argues that a different type of concept – a gradience analysis – is needed here in order to explain the differences between ‘create’ and ‘produce.’

In order to seek more information about the types of objects, the most frequent (top ten) object types from the Brown Corpus and the Frown Corpus were examined and are shown in Table 4 below. A comparison to a larger corpus, the British National Corpus (BNC), was also undertaken (see Table 5 to follow).

Table 4: Top Ten Objects of ‘Create’ and ‘Produce’ in the Brown Corpus and the Frown Corpus

‘Create’			‘Produce’		
Objects	Frequency	%	Objects	Frequency	%
<i>jobs</i>	15	3.21	<i>effect</i>	14	2.99
<i>world</i>	9	1.93	<i>results</i>	12	2.56
<i>problem</i>	9	1.93	<i>change</i>	7	1.50
<i>opportunities</i>	8	1.71	<i>response</i>	6	1.28
<i>image</i>	6	1.28	<i>energies</i>	5	1.07
<i>thing</i>	5	1.07	<i>power</i>	4	0.85
<i>situation</i>	5	1.07	<i>energy</i>	4	0.85
<i>environment</i>	5	1.07	<i>variables</i>	3	0.64
<i>effect</i>	5	1.07	<i>forms</i>	3	0.64
<i>right</i>	4	0.86	<i>evidence</i>	3	0.64

From Table 4, one can see that the most frequently appearing objects for both verbs are mostly abstract. ‘Jobs’ is most often seen with ‘create,’ and ‘effect’ is most often seen with ‘produce.’ Referring back to Table 2, most of these objects fall under the ‘bring into existence’ sense for ‘create’ and ‘cause to exist’ for ‘produce,’ indicating that a majority of the top objects are constituted by these two senses which are identical.

The collocate lists in Table 5 were collected through the English Sketch Engine (Kilgarriff and Tugwell, 2001). With the help of the Sketch-Difference function, we listed only the collocates that were found with one of the verbs (i.e., they do not appear with the compared verb). The frequency indicates the frequency of the verb-object combinations.

Table 5: Types of Objects That Appear with ‘Create’ and ‘Produce,’ Respectively (only the top results are presented)

‘Create’-only		‘Produce’-only	
Objects	Freq.	Objects	Freq.
<i>job</i>	491	<i>result</i>	515
<i>problem</i>	395	<i>report</i>	319
<i>atmosphere</i>	208	<i>goods</i>	188
<i>condition</i>	197	<i>evidence</i>	187
<i>opportunity</i>	174	<i>document</i>	175
<i>impression</i>	165	<i>material</i>	172
<i>environment</i>	154	<i>product</i>	157
<i>image</i>	144	<i>book</i>	146
<i>situation</i>	133	<i>figure</i>	137
<i>system</i>	133	<i>output</i>	136
<i>market</i>	129	<i>range</i>	117
<i>world</i>	123	<i>list</i>	101
Modifier		Subject	
<i>newly</i>	162	<i>company</i>	109
		Modifiers	
		<i>only</i>	179
		<i>now</i>	107

Based on Tables 4 and 5, it was found that the top few collocates from the two corpora (Table 4) consist mostly of abstract objects for both verbs. A larger sample from the British National Corpus shows that more concrete nouns (bold) were found with ‘produce’ than with ‘create.’

Based on the object arguments of both verbs, we found distinctive differences, which are important for distinguishing the two verbs. For instance, we discovered that ‘create’ is usually followed by objects that are more often abstract in nature, while ‘produce’ is usually followed by objects that are more concrete and that might represent massive quantities, such as objects that are manufactured in factories (e.g., ‘machines,’ ‘seeds,’ ‘goods,’ etc.).

Because ‘produce’ usually denotes factory-made products, the objects it collocates with are naturally fixed, disallowing much creativity. In contrast, ‘create’ seems to be followed by objects whose properties are not fixed (such as ‘problems’ and ‘image’) and allow more creativity (usually artistic-related). These two differences may explain our earlier example in (6b), whereby ‘*to over-create’ is implausible while ‘to over-produce’ seems to make more sense. The analyses enable us to explain why certain constructions are allowed or disallowed with a certain verb.

4. Discussion

From previous analysis, the WordNet senses were listed as separate entries. For instance, the meanings of ‘create by artistic means’ and ‘create or manufacture a man-made product’ (cf. (8) earlier) were not explicitly linked in WordNet. In this section, we will show that the seemingly discrete senses from a dictionary (WordNet in this case) can be re-analyzed as in Figure 1 below. Figure 1 also provides the corresponding senses taken from Table 2 in square brackets.

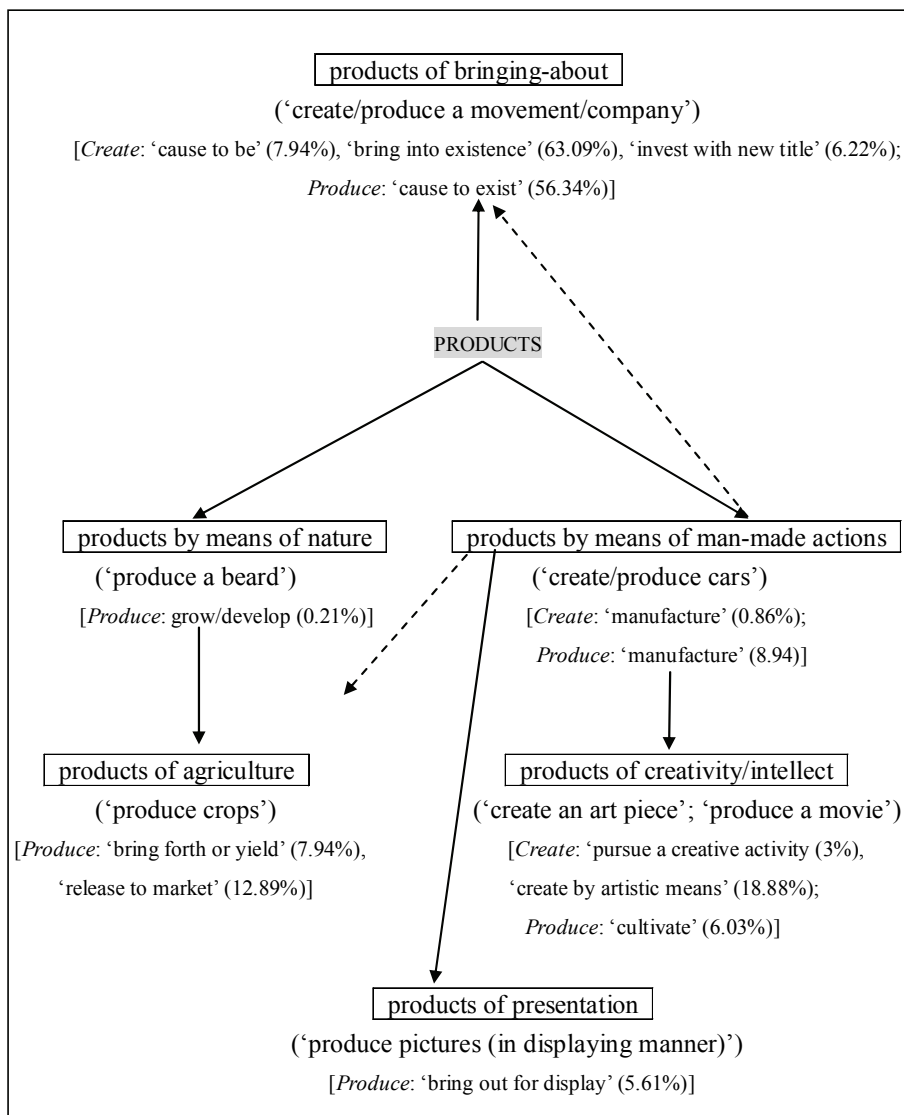


Figure 1: Re-analyzing Senses of 'Create' and 'Produce'

In Figure 1, one can see that all the meanings of 'create' and 'produce' can be derived from the core meaning of PRODUCTS. The distributional data from the corpora are also seen. For instance, one can see the prominence of metaphorical meanings for PRODUCTS in 'create' and 'produce' among all other meanings—A majority of the corpora instances come from the more metaphorical meaning of PRODUCTS ('products

of bringing about’). An example is ‘create/produce a movement or a company’ in which ‘movement’ and ‘company’ are not physical products of the action. They refer to notions that are more abstract. This high occurrence of metaphorical use can also explain the results from the top collocated objects in the previous section, in which the objects of both verbs are constituted mostly by abstract objects. Distributional information as such cannot be obtained without the use of a corpus. Conversely, PRODUCTS under ‘products by means of nature’ and ‘products by means of man-made actions’ usually denote products that are more physical. Under ‘nature,’ there is ‘products of agriculture.’ In this sense, only the use of ‘produce’ was found.

As agricultural products cannot be separated from human labor, there is a dotted link between ‘products of agriculture’ and ‘products by means of man-made actions.’ Similarly, humans cause the rise of events, such as a movement or the building of an organization. Therefore, there is also a dotted line linking ‘man-made actions’ to ‘products of bringing-about.’ Under ‘products of man-made actions,’ there are ‘products of creativity/intellect’ and ‘products of presentation,’ of which the latter denotes a physical action in presenting something, usually referring to proof or evidence, in a straightforward manner.

The meanings of PRODUCTS in Figure 1 are the result of the re-analyses of WordNet senses. By doing so, this paper argues that the seemingly unrelated meanings, as listed in WordNet, could be related at a higher, conceptual level. From Figure 1, one can see the ‘produce’-only senses – ‘products by means of nature,’ ‘products of agriculture,’ and ‘products of presentation.’ From information as such, we can obtain hints regarding the overlapping use in ‘products of bringing-about,’ ‘products by means of man-made actions,’ and ‘products of creativity/intellect.’ (Note that the similarity in ‘products of creativity/intellect’ cannot be obtained by using the sense listing in Table 2.) Yet, however inter-related the senses are in Figure 1, it is unable to display precisely the differences between ‘create’ and ‘produce.’ In order to make the results complete, we hypothesized that a gradience analysis must be undertaken, which is elaborated below.

A gradience analysis of meanings has been adopted in many studies – as early as 1961 by Bolinger. Coleman and Kay (1981) are among the scholars who related gradience in linguistics to prototypical semantics. In addition, there are also studies that locate gradience in grammar. Owen (2009), for instance, used gradience analysis in

analyzing the headedness of *Noun1 of Noun2* ('a lot of money,' 'a hatred of money,' 'a child of money,' etc.), in which headedness was found shifting from *Noun1* to *Noun2* gradually. Sorace and Keller (2004) had also used this notion to examine grammaticality judgments by native speakers. Furthermore, a representation of near-synonyms using hierarchies can also be seen in the computational model by Edmonds and Hirst (2002), which compared near-synonyms using a representation of word meaning through different levels of granularity. Words at a more fine-grained level were gathered under coarse-grained ontological concepts. The fine-grained level comprises attributes or peripheral concepts with various degrees of meaning. For instance, the word 'error' was distinguished from 'blunder' by being medium in 'Blameworthiness' and low in 'Concreteness,' whereas 'blunder' was found to be comparatively high in both 'Blameworthiness' and 'Concreteness.'

In this section, we explain the similarities and differences between 'create' and 'produce' based on a gradience principle, which serves as a complement to our previous sense analyses in Figure 1. Our representation is given in Figure 2 below.

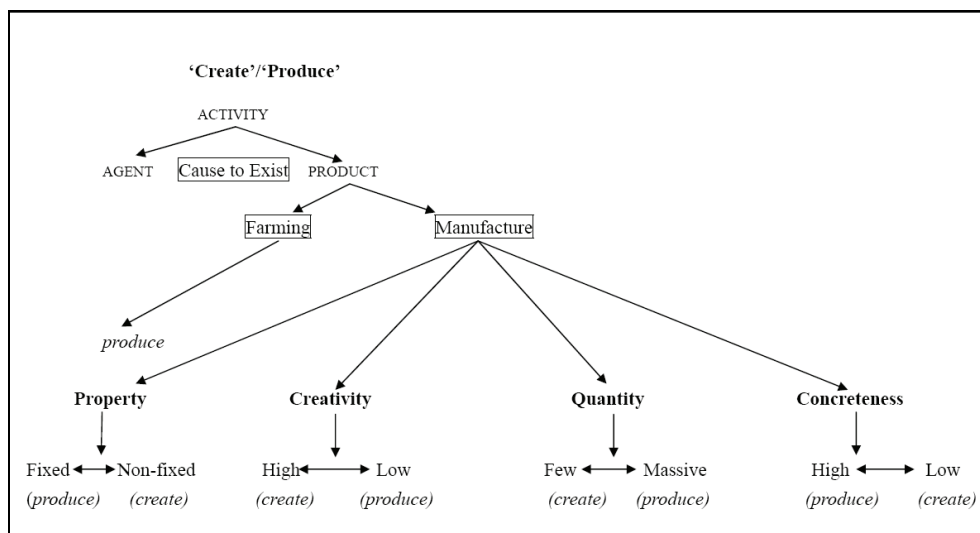


Figure 2: Core and Peripheral Meanings of 'Create' and 'Produce'

Our representation in Figure 2 allows the display of eventive information encompassing an ACTIVITY, an AGENT, and a PRODUCT. An AGENT (which can be a

person or an organization) will ‘bring about the existence’ of a PRODUCT through an ACTIVITY. ‘Farming,’ one of the distinguishing features found from the analysis of WordNet senses and from the ‘produce’-only example in ‘products by means of nature’ and ‘products of agriculture’ in Figure 1, is a specific PRODUCT used only with ‘produce’ (e.g., ‘their seeds **produce** vigorous blooming plants’). This explains the node of ‘Farming’ in Figure 2.

Another shared PRODUCT is ‘Manufacture,’ which comprises different types of PRODUCTS distinguishable through an inspection of scaled information in semantic features, such as ‘property,’ ‘creativity,’ ‘quantity,’ and ‘concreteness,’ represented as antecedents of ‘Manufacture’ in Figure 2 above. These semantic features cannot be shown in Figure 1 by inspecting the senses and their relations per se. Therefore, Figure 2 can be seen as an in-depth analysis of the meanings discovered previously – that is, even though some senses are shared by ‘create’ and ‘produce,’ their distinctions can only be made clearer with additional interpretation from Figure 2.

In Figure 2, the verbs falling at both ends of the scales are given in brackets. The ends of the continuum show a tendency toward using a particular verb. For instance, if a product has high concreteness and is manufactured in massive amounts, as in (12a) below, there is a high possibility that the verb is ‘produce’ rather than ‘create.’ However, if the product is concrete yet the level of creativity cannot be stated clearly from the context, as in (12b) below, both ‘produce’ and ‘create’ are possible (but a different reading is expected). When ‘he creates smoking guns,’ he is more likely to design a first sketch of the smoking guns. In contrast, when he ‘produces smoking guns,’ he produces in a large amount. (‘Produce’ in (12b) can also mean ‘display all the smoking guns as a proof’.)

- (12) (a) *We have been able to _____ and deliver our machines throughout 1960.* (produce)
 (b) *He cannot _____ the smoking guns they demand.* (produce/create)
 (c) *Before he was forty, Griffith had _____ the art of the film.* (created)

On the other hand, if a product is high in ‘creativity,’ the verb ‘create’ is likely to be used, as in (12c). Therefore, from (12a) through (12c), a gradient decrease in

creativity and in quantity is observed. However, sometimes it is not easy to separate the semantic features in the examples, as no example can be purely ‘creative’ but neither ‘concrete’ nor ‘abstract,’ etc. In (13) below, we show the objects that come in different quantity.

- (13) (a) *The Holocaust* _____ *a world of its own.* (created)
(b) *Jobs in the U.S., Asia, and Europe are* _____ *by smaller companies.*
(created)
(c) *a bleak motif with barren octaves* _____ *a rather ancient effect*
(creating/producing)
(d) *because the Bahia region* _____ *most of the world's cacao beans*
(produced)
(e) *Polyether foams on the market had to be* _____ *by the two-step prepolymer method.* (produced)

In (13a), with ‘a world’ (in which there is only one), the verb is always ‘create.’ When the noun represents many, as in (13b) (‘jobs’), ‘create’ is still preferred. When the noun becomes less certain in quantity, which sometimes might appear in singular or plural forms (‘effect’ or ‘effects’) as in (13c), both ‘create’ and ‘produce’ are seen. When the noun is massive and is usually ‘multiplied,’ especially as agricultural products, ‘produce’ is the preferred verb, as in (13d). In (13e), when the noun is massive and might be uncountable, ‘produce’ is usually used. Similarly, moving in a reverse direction from (13e) to (13a), one also sees a gradient change from a more definable property to a less definable one.

The examples in (13) above show how a gradience analysis might explain the (non-)overlapped meanings between the two verbs. From these examples, one can observe the use of these scaled semantic features in determining the selection of ‘create’ and ‘produce.’ Based on the above, we found that differentiating ‘create’ and ‘produce’ allows a tendency reading of a word through which a prediction can be made for the choice of verb to select. For instance, given a definition of a verb that denotes an activity between an AGENT and a PRODUCT, in which the product is manufactured in massive quantities, the verb suggested is likely to be ‘produce’ rather than ‘create.’ A

model as such can be further utilized for computational purposes, especially in predicting a verb choice given its surrounding features. As for the case of ‘create’ and ‘produce,’ although continuity is found to be a better way to explain their differences, “[o]ur task is not to choose among them [continuity or discreteness], but rather to explicate the specific ways in which this fundamental opposition plays itself out across the full range of linguistically relevant phenomena.” Highlighting the importance of continuity in semantics is by no means demoting the role of discreteness (a statement based on Langacker (1994: 18)).

5. Conclusion

This work proposes a corpus-based inspection of two verbs using corpus linguistic techniques, as well as a gradience analysis of similarities and differences of two near-synonyms. Two questions were presented in this paper. The first question was related to discerning the similarities and differences between ‘create’ and ‘produce,’ while the second question focused on whether the results from a corpus-based methodology could be represented in any forms in addition to displaying the percentages from the corpora. The answer we provided for each question was positive.

The corpus provided quantitative linguistic data for analyses but, as demonstrated in this work, we also incorporated a gradience analysis, in addition to re-structuring the conceptual relatedness of meanings for both verbs. We also realized that although many of the senses of the two verbs, and even their object types, might seem unrelated, they could be linked through the notion of PRODUCT (as products of agriculture, as products of creativity, as products of presentation, etc.). We also provided an explanation encompassing the non-discrete semantic features of ‘create’ and ‘produce’ and discussed the literal and/or metaphorical extensions of PRODUCTS of both verbs. Senses that seemed to be unrelated were re-structured and represented as a conceptual system. Similarities within several shared senses were distinguished using scaled-based semantic features.

Regarding the similarities and differences between the verbs ‘create’ and ‘produce,’ our results showed the following findings. First, ‘create’ and ‘produce’ have two similarities: (a) they both appear most frequently as a bare infinitive and in the *-ed*

forms; and (b) they share two overlapping senses, namely ‘bring into existence/cause to happen, occur, or exist’ and ‘create or manufacture a man-made product.’ Information about verb forms and senses are sometimes available in dictionaries but only a corpus will provide distributional information. Second, in terms of selectional restriction, there are preferences for the selection of object argument for ‘create’ and ‘produce’ – ‘create’ tends to be followed more often by abstract objects than ‘produce’ does. ‘Create’ is usually followed by objects that are fewer in quantity and whose properties are not fixed, and these objects therefore possess a higher possibility of applying creativity in its creation. Comparatively, ‘produce’ is usually followed by objects that are produced in great number (usually manufactured by factories), with a low threshold for creativity. Third, only ‘produce’ was found to occur in the sense of ‘Farming.’ Most of the semantic differences (i.e., ‘property,’ ‘creativity,’ ‘quantity,’ and ‘concreteness’) are better measured on a scale rather than in discrete dichotomies.

Our suggestion of a hybrid model including sense analyses and hierarchical representations was found most adequate to capturing the similarities and differences between the two verbs. In addition, we also argued for the inter-relatedness of meanings for the seemingly discrete senses provided in traditional dictionaries. This was achieved through re-analyzing the senses from WordNet, by adding distributional information from the corpora.

For future work, more near-synonyms can be attested using the model suggested herein. The similar or differing translation for ‘create’ and ‘produce’ into Mandarin can also be considered for cross-linguistic work. Mandarin work on creation-related verbs can be seen in Hung (2010). Furthermore, incorporated ontologies based on more than one language will also provide additional cross-linguistic references for near-synonym research. It is also possible to examine predictability across languages based on closely related words.²

On the other hand, we also found that our analysis is, to some extent, in accordance with Pustejovsky’s (1991) notion of coercion, in which the semantic load of

² It is by coincidence that in the author’s recent work (Chung, 2010) on the investigation of the classifier *buah* in Malay, an Austronesian language, the nouns classified by this classifier are all types of products. The author is interested in finding out whether the verbs preceding this classifier are also verbs of creation/production.

the nominals may be shared by the verbs. For instance, the metaphorical reading in ‘to create a market’ may come from the ‘market,’ whereas the literal reading in ‘to create a healthy two-system eastern railroad’ may come from the physical image of the ‘railroad.’ Pustejovsky’s (1991: 426-427; 1995) further proposal of qualia structure in the nominals following the four roles in (14) below has also received similar research on verbal semantics.

- (14) (a) Constitutive role (i.e., the relation between an object and its proper parts, such as ‘narrative’ for a novel)
- (b) Formal role (i.e., role that distinguishes the object within a larger domain, such as ‘book’ or ‘disk’ for a novel)
- (c) Telic role (i.e., the purpose and function of the object, such as ‘read’ for a novel)
- (d) Agentive role (i.e., factors involved in bringing about the object, such as ‘artifact’ or ‘write’ for a novel)

The four roles in (14) were found to be able to distinguish meanings of polysemy, too. For instance, ‘a red pen’ can carry both the formal role (‘a red pen, not a blue pen’) and the telic role (‘a pen with red ink, not a pen that is red’) (cf. Croft and Cruse, 2004: 137). In the examples above, ‘market’ in ‘to create a market’ emphasizes the ‘offer and demand’ of transactions in the market (thus, a telic role). On the other hand, ‘railway’ in ‘to create a railway’ emphasizes the agentive role, whereby the bringing-about of the artifact is highlighted. In (15a) below, the distinction of ‘Mrs. Gaskell’ as a person or as a character in the novel can be seen as the formal role in which the role based on a larger domain is emphasized. Such an example can be explained using a metonymy model; however, this is not the focus of the current paper.

- (15) (a) *she denied in **creating** ‘Mrs. Gaskell’*
- (b) *[the] family of Luxemburg immigrants, in fact, **produced** two exceptional children*
- (c) *biological beings, family men, those who **produce** the next generation*

In (15b) and (15c), the relation between an object and its proper parts, such as ‘children’ (15b) and ‘generation’ (15c), is shown for ‘produce.’ However, this part of the analysis still requires more in-depth investigation because the roles in (14) are noun-based and therefore, when applied to verbs, each individual noun must be inspected separately. We will reserve this topic for future research.

REFERENCES

- Anthony, L. (2005), “AntConc: Design and Development of a Freeware Corpus Analysis Toolkit for the Technical Writing Classroom,” in the *Proceedings of the International Professional Communication Conference*, (IPCC), 729-737.
- Biber, D., S. Conrad, and R. Reppen. (1998), *Corpus Linguistics: Investigating Language Structure and Use*, Cambridge: Cambridge University Press.
- Bolinger, D. L. (1961), *Generality, Gradience, and the All-Or-None*, The Hague: Mouton.
- Chung, S.-F. and K. Ahrens (2008), “MARVS Revisited: Incorporating Sense Distribution and Mutual Information into Near-Synonym Analyses,” *Language and Linguistics: Lexicon, Grammar and Natural Language Processing*, 9(2), 415-434.
- Chung, S.-F. (2010), “Numeral Classifier Buah in Malay: A Corpus-based Study,” *Language and Linguistics*, 11(3), 553-577.
- Church, K. W., W. Gale, P. Hanks, D. Hindle, and R. Moon. (1994), “Lexical Substitutability,” in B. T. S. Atkins and A. Zampolli (eds.), *Computational Approaches to the Lexicon*, 153-177.
- Coleman, L. and P. Kay (1981), “Prototype Semantics: The English Word Lie,” *Language*, 57(1), 26-44.
- Collinson, W. E. (1939), “Comparative Synonymics,” *Transactions of the Philosophical Society*, 54-77.
- Croft, W. and D. A. Cruse (2004), *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- Cruse, D. A. (1986), *Lexical Semantics*, Cambridge: Cambridge University Press.
- Dixon, R. M. W. (1991), *A New Approach to English Grammar; on Semantic Principles*, New York: Oxford University Press.
- Dowty, D. (1979), *Word Meaning and Montague Grammar*, Dordrecht: Reidel.
- Dowty, D. (1991), “Thematic Proto-Roles and Argument Selection,” *Language*, 67(3), 547-619.
- Edmonds, P. and G. Hirst. (2002), “Near-Synonym and Lexical Choice,” *Computational Linguistics*, 28(2), 105-144.
- Fellbaum, C. (1998), *WordNet: An Electronic Lexical Database*, MIT Press.
- Harris, R. (1973), *Synonym and Linguistic Analysis*, Oxford: Basil Blackwell.

- Heine, B. (1997), *Cognitive Foundations of Grammar*, New York: Oxford University Press.
- Hung, W.-T. (2010), "Analyze the near-synonyms 'Manufacture' and 'Produce' ," *Journal of Applied Chinese*, 6, 223-246.
- Kilgarriff, A. and D. Tugwell (2001), "WORD SKETCH: Extraction and Display of Significant Collocations for Lexicography," *ACL Workshop COLLOCATION: Computational Extraction, Analysis and Exploitation*, 32-38, Toulouse.
- Landes, S., C. Leacock, and R. Teng. (1998), "Building Semantic Concordances," in C. Fellbaum (ed.), *WordNet: An Electronic Lexical Database*, 199-216, MIT Press.
- Langacker, R. D. (1994), "The Limits of Continuity: Discreteness in Cognitive Semantics," in C. Fuchs and B. Victorri (eds.), *Continuity in Linguistic Semantics, Lingusticæ Investigationes: Supplementa Series*, 9-20, Amsterdam/Philadelphia: John Benjamins.
- Leibniz, G. W. von. (1903), "Table of Definitions," in L. Couturat (ed.), *Opuscles et fragments inédits de Leibniz*, Paris.
- Lyons, J. (1995), *Linguistic Semantics: An Introduction*, Cambridge University Press.
- MacLaury, R. E. (1997), *Color and Cognition in Mesoamerica: Constructing Categories as Vantages*, Austin: University of Texas Press.
- MacLaury, R. E. (2002), "Introducing Vantage Theory," *Language Sciences*, 24, 493-536.
- Parsons, T. (1990), *Events in the Semantics of English*, Massachusetts: The MIT Press.
- Pustejovsky, J. (1991), "The Generative Lexicon," *Computational Linguistics*, 17(4), 409-441.
- Pustejovsky, J. (1995), *The Generative Lexicon*, Cambridge, Massachusetts, and London: The MIT Press.
- Sorace, A. and F. Keller. (2004), "Gradience in linguistic data," *Lingua*, 115(11), 1497-1524.
- Taylor, J. R. (1995), *Linguistic Categorization: Prototypes in Linguistic Theory*, Oxford: Oxford University Press.
- Taylor, J. R. (2002), "Near Synonyms as Co-extensive Categories: 'High' and 'Tall' Revisited," *Language Sciences*, 25, 263-284.
- Xiao, R. and T. McEnery. (2006), "Collocation, Semantic Prosody, and Near Synonym: A Cross-Linguistic Perspective," *Applied Linguistics*, 27(1), 103-129.

以語料庫爲本分析近義詞 **Create** 和 **Produce**

鍾曉芳*

摘 要

本文旨在以語料庫研究 create 和 produce 此組近義詞，得出此組近義詞與 product 相關之事件類型搭配，能夠說明兩者間的異同。語料來自 ICAME 所收錄之 Brown Corpus 和 Frown Corpus 兩語料庫，並與 British National Corpus 中的 Sketch Engine 搭配詞搜尋法呈現的語料做比較。語料庫分析結果發現，有些 product 事件(如：Manufacture)均能與兩個動詞搭配；而其他 product 事件(如：Farming)只能與其中一個動詞(Produce)搭配使用。此外，利用梯度分析法(Gradience Analysis)，本文以事件的本質(Property)、創造力(Creativity)、數量(Quantity)及其具體性(Concreteness)等語意成分區分 create 和 produce 的篩選機制，並以層級結構呈現 create 和 produce 的異同處。

關鍵詞：近義詞、語料庫、Create、Produce、語義分析、梯度分析法

* 作者爲國立政治大學英國語文學系助理教授，E-mail: sfchung@nccu.edu.tw。