Hitzen ondoeza izena eman zion Joseba Sarrionandiak hitzen esanahi bikoitzak bilatzen zituen hiztegi bati. *Hitzokei* deitu zion norbaitek antzeko ariketa kolektiboari sarean. Hitzekin jolasean aritzea giza jolasa da, eta ez dakit makinek egiteko gaitasuna izango duten noizbait. Hau da, makina batek txisteak egin ditzake, baina ez dakit ulertzera iritsiko den txiste horren muina zein den, zergatik egiten dugun barre, zergatik den gozagarri (edo umorearen limiteen eztabaida irekita dagoen honetan, zergatik den amorrarazle).

Hitzek ez dute esanahi bat eta bakarra. Labur adierazteko, *Euskadi ta Askatasuna*ko eta *Libertad Digital*eko "*askatasuna*" hitzak ez dute esanahi berdina. "Gipuzkoako ingurumen guneak" deitzen zaio kutsatzeko helburua duen erraustegi bati, "prozesu parte-hartzaile" deitzen zaio parterik hartu ezin daitekeen zerbaiti. Adibide asko eman daitezke Orwellek 1984 liburuan "*doblepentsatu*" deitzen zaion horri buruz hitz egiteko, errealitatea eta hitzak bat ez datozen kasu horiek gogora ekartzeko ("sumar para ganar", "lehen, orain eta beti, aberria", "herritarra helburu", "sí es sí" edo berriki Pedro Sánchezek defendatutako errepublikanismo juankarlista eta plurinazionaltasun monoburujabea kasu).

Gizakiok gai gara hitz horien atzean dauden geruzak ulertzeko, eta singularitatea noizbait etorriko bada, makinak ere gai izan behar dira hitzen atzeko esanahiak ezagutzeko.

Wikimedia Fundazioa 2030rako estrategia diseinatzen ari da. Eztabaida interesgarria egon da 2030ean inteligentzia artifizialak eta "deep learning" famatu horrek ekarriko digunaren inguruan. Milioika artikulu izan ditzakegu 2030an euskaraz, informazio basikoarekin, baina… beti beharko al dugu gizakion eskua artikulu horiek gizakiontzat baliagarri izateko? Adibide batekin hornitzen zuen ingelesezko Wikipediako lankide batek: "*gizaki guztiak berdinak dira. Makina batentzat esaldi ona da, gizakiontzat ez du zentzurik*". Izan ere, hor sartzen da makina eta pertsonen arteko alderik nabarmenena: esaldi bat ("Gizaki guztiak berdinak dira") testuinguruan jartzen jakitea (barka bezate Mike Godwin jaunak eta Gandhiren biktima hegoafrikar ez-zuzenekoek: "*Bai, baina Hitler ez zen Gandhi*").

Buelta gaitezen hizkuntzara. Facebook publizitate enpresak negoziatzeko gaitasuna duen inteligentzia artifizialeko bot bat sortzeko helburua du. Publizitatea hobeto saltzeko esfortzua izango da, ziur. Bot horrek lortu du tratu-regateo klasikoan sartzen diren estrategiak bilatzea: nahi duzun objektuari arretarik ez jartzea eta inguruko besteen ingurua negoziatzen hastea, akordiorik onena lortzeko enpatia lantzea, eskaintza txar bat hartzea inongo eskaintza baino hobeto… Labur azalduta, honela egiten du: item kopuru bat saldu-erostea da helburua, eta item bakoitzak puntu kopuru batzuk ditu. Negoziazioaren bukaeran transakzio osoari puntuak esleitzen dizkio, eta atzerantz egiten du, erabili dituen hitz, esamolde eta estrategiak puntuatzen. Geroz eta puntu gehiago lortu, orduan eta "ohikoagoa" bilakatuko du estrategia bat. Suposa dezakegu erosleari "putakume" deitzeak ez duela puntuatzen, baina hitz politak esateak baietz. Suposa dezakegu.

Eta zer harreman du honek hizkuntzarekin? Beno, ba Facebookek bi bot jarri ditu elkarren artean negoziatzen, estrategiak fintze aldera. Eta hara non, euren artean, ikerlarientzat ezezaguna zen hizkuntza bat garatu duten. Roboten negoziazio hizkuntza. Hizkuntza hau erabiltzeak puntu gehiago emango zizkion programari, noski. Robotei, beraz, hizkuntza berririk EZ asmatzeko araua sartu behar zaie, eta euren hizkera kontrolatu.

Buelta gaitezen Wikimediara. Beste batzuetan hitz egin dugu Wikidatari buruz. Laburbilduz:

Wikidata datuak biltzeko ekimen libre erraldoia da. Bertan daude Wikipediako elementuen arteko harremanak, zein OpenStreetMapek elementuen datuk, baita euren arteko loturak ere. Kanpo datubaseak ere lotzen dira eta orain gutxi hasi dira ere zientzia argitalpenen bildumak egiten, metadatu guztiak bertan sartuta.

Wikimedia Fundazioak ere Wikhiztegia barnebiltzen du. Wiktionary izenarekin, inoiz egin den meta-hiztegirik handiena da. Azalduko dizuet labur: bi hiztegi mota izan ditugu orain arte. Lehenengo mota euskara-euskara hiztegiak ziren, hau da, hitz bat bilatu eta bere esanahia zutenak. Bigarren mota euskara-ingelesa edo errusiera-arabiera hiztegiak dira: hitz bat bilatu eta beste hizkuntza batean nola esaten den adierazten duten hiztegiak. Wikhiztegiak biak barneratzen ditu, baina beste optika batetik. Euskara-euskara hiztegia izan daiteke, ingelesa-euskara hiztegia ere (edo munduko beste edozein hizkuntza) eta, aldi berean, ingelesezko/errusierazko/abkhazierazko… hitz horren esanahia esaten digu gure hizkuntzan. Eta hori gutxi balitz, hiztegi etimologikoa ere bada, hizkuntzen arteko harremanak eta jatorriak emanez. Adibide interesgarri batzuk: ***oso***, ***alaba***, ***pago***, ***osaba***, ***ardo…*** konturatuko zinetenez hitz guzti hauek ***“Oso [h]ondo printzipioa”*** betetzen dute: euskaraz “*muy bien*”, gaztelaniaz “*hartz sakona*”. Euskarazko Wikhiztgia nahiko ahul dago, baina baliteke hurrengo urteotan loraldia izatea, Wikidatari esker.

Wikidata, esan bezala, datu base erraldoia da. Hurrengo hamarkada osoan zehar Wikhiztegia (oraingoa zein hamarkada berrian sortuko dena) barneratzeari ekingo zaio, “lexema” guztiak gordez. Lexema bakoitzaren hizkuntza, forma gramatikala, kategoria lexikala, esanahia(k), errimak, erabilera adibideak eta itzulpenak gordeko dira. Hau da, hamar bat urte iraun dezakeen proiektu honek munduko ahalik eta hizkuntza gehienetako hitz guztiak erregistratu, katalogatu, lotu eta analizatu nahi ditu, itzultzaile automatikoak hobetuz eta etorkizuneko Wikipedian itzulpenen zentzua aldatuz.

Buelta gaitezen hasierara. Facebookek lortu du bi robotek, nahi gabe, hizkuntza berri bat asmatzea. Google eta beste mega-enpresak ere ari dira esfortzu berdinean, inteligentzia artifizialaren eremuan. Aldi berean, librea izango den hizkuntza guztietako hiztegi erraldoi baten datu-basea eraikiko du komunitateak, nahi duten robotek nahi duten hizkuntza ulertu eta hitz egiteko aukera izan dezaten. Hizkuntzaren prozesamendu naturala errealitate bat da, eta asko hedatuko da hurrengo bi hamarkadatan zehar. Bai, baina jakingo al dute robot horiek hitzen ondoeza ulertzen? Jakingo al dute “askatasuna” hitzak esanahi ezberdinak dituela hartzailearen arabera? Jakingo al du etorkizuneko robot horrek publizitatea, propaganda edo ekintza politikoa ez dela literala?

Zabaldu:

- Click to share on Twitter (Opens in new window)
- Click to share on Facebook (Opens in new window)
- Click to share on Telegram (Opens in new window)
- Click to share on Pocket (Opens in new window)